ASPECTS OF LITHOAUTOTROPHY IN IRON-OXIDIZNG THERMOACIDOPHILIC ARCHAEA

by

Timothy David Williams (B.Sc.)

This thesis is presented for the
Degree of Doctor of Philosophy,
in the Department of Biological Sciences,
University of Warwick.

September, 1995.

Contents

List of Figures

Acknowledgements

I would like to thank my supervisor, Dr. Paul Norris, for his enthusiasm and constructive input throughout the course of this project.

Thanks also to my father and stepmother, for whose continual support I am most grateful.

I would like to thank all members of the Microbiology I research group for much help and advice, especially John, Darren, Nick, Sara, Martin, Jo and Howard. Thanks also to Kirsty, Dave and Heather. I would also like to remember Jim Trickett, sadly missed by all who knew him.

I acknowledge the support of the BBSRC. in funding this studentship.

Declaration

I declare that this thesis has been composed by myself and has not been used in any previous application for a degree. All results presented within have been obtained by myself under the supervision of Dr. Paul Norris.

Abbreviations

A	absorbance
ATP	adenosine triphosphate
BC	biotin carboxylase
BCCP	biotin carboxyl carrier protein
CoA	coenzyme A
Da	Dalton
DIG	digoxigenin
DMB	3,3'-dimethoxybenzidine
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylene(diamino)tetra acetate
FPLC	fast protein liquid chromatography
IEF	isoelectric focussing
kb	kilobase
NAD	nicotinamide adenine dinucleotide
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
Pa	Pascal
PEG	polyethylene glycol
PEP	phosphoenolpyruvate
pers. comm.	personal communication
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
RuBisCo	ribulose bisphosphate carboxylase/oxygenase
SDS	sodium dodecyl sulphate
SSC	sodium chloride / sodium citrate buffer
TBE	tris / borate / EDTA buffer

TCA	tricarboxylic acid
TE	tris / EDTA buffer
TEMED	N,N,N',N'-tetramethylethylenediamine
TMBZ	3,3',5,5'-tetramethylbenzidine
UV	ultra violet
v/v	volume per volume
w/v	weight per volume

Abstract

The phylogeny and proteins of chemolithoautotrophic growth of thermoacidophilic *Sulfolobus*-like archaea were investigated.

Sulfolobus-like strain HT is a high growth temperature chemolithotrophic isolate with potential application in mineral sulphide bioleaching. The gene encoding its 16S rRNA was cloned and sequenced. Phylogenetic analysis of this sequence showed strain HT segregating within the genus *Sulfolobus*. It showed greater similarity to the sequence of *Sulfolobus shibatae*, a facultative chemolithotroph, than to the sequence of *Sulfolobus acidocaldarius*, an obligate heterotroph. Flanking regions of the 16S rRNA gene were also sequenced, showing secondary structure similarity to those of *S. acidocaldarius*, implying a similar excision and processing pathway.

A protein of 330 kDa, consisting of 59 kDa and 19 kDa subunits, was over-expressed during CO_2-limited autotrophic growth of *Sulfolobus* strain LM and had previously been shown to co-purify with ATP and acetyl-CoA dependent CO_2 uptake. The 59 kDa subunit was partially purified and its N-terminal amino acid sequence obtained. The gene encoding this polypeptide was cloned and sequenced. An open reading frame likely to encode the 19 kDa subunit was adjacent to this gene, forming a possible operon. Homology searches revealed that the predicted amino acid sequence of the 59 kDa subunit was similar to those of ATP-dependent biotin carboxylase enzymes, predicted active site residues being conserved. Homology searches of the predicted amino acid sequence of the ORF likely to encode the 19 kDa subunit revealed similarity to biotin carboxyl carrier proteins, with a biotin binding motif being conserved.

In *Sulfolobus* LM, a polypeptide of 27 kDa molecular weight was over-expressed during autotrophic growth on ferrous iron in comparison with autotrophic growth on tetrathionate. This polypeptide was partially purified and its N-terminal amino acid sequence obtained. After the cloning and sequencing of the gene encoding this protein by a co-worker, homology searches were carried out. It showed homology to the alkyl hydroperoxide reductase / thiol specific antioxidant (AhpC/TSA) family of proteins, members of which are thought to play a role in protection against oxidative stress. The predicted amino acid sequence was phylogenetically analysed, segregating within a group of sequences derived from eukaryotes and archaea, which possess one conserved cysteine residue, as opposed to a group consisting of eukaryotes and eubacteria, possessing two conserved cysteine residues.

A membrane bound cytochrome showing a difference spectrum alpha absorbance peak at 572 nm had previously been found to be present only during ferrous iron oxidation in thermoacidophilic archaea. This novel cytochrome was partially purified from the membrane fraction of *Sulfolobus* strain LM autotrophically grown on ferrous iron. It was shown to retain haem staining activity after SDS treatment, thus allowing its identification as a polypeptide of approximately 66 kDa. A procedure which may allow the N-terminal sequencing of this protein and the initiation of its molecular biological study was identified.

CHAPTER 1

INTRODUCTION

1.1 - Project aims

The object of this study was to initiate a molecular biological approach to the elucidation of mechanisms of chemolithoautotrophic growth of ferrous iron- and sulphur-oxidising thermoacidophilic archaea. In addition, the future phylogenetic classification of the principal organisms under study was to be facilitated by 16S rDNA sequencing. The initiation of these studies was considered important for several reasons.

First, the thermophilic crenarcheota are of great potential industrial use, particularly in mineral sulphide processing. The consideration of various strains for commercial application would be facilitated by their reliable identification and classification and elimination of confusion in earlier species designations. In addition, the elucidation of the key metabolic processes involved in growth on mineral sulphides should aid in process optimisation and, in the longer term, in strain improvement.

Secondly, the previously unstudied mechanisms of chemolithoautotrophic growth would be expected to involve novel proteins and metabolic pathways. These are of intrinsic interest and should also provide insights into aspects of thermostability, stress tolerances and molecular evolution.

The specific areas selected for study were the mechanisms by which *Sulfolobus*-like organisms oxidize ferrous iron and assimilate carbon dioxide. The essential procedures were to be the study of proteins over-expressed during ferrous iron oxidation and during growth under CO_2 limitation. The organisms to be used were *Sulfolobus* strain LM and a *Sulfolobus*-like organism capable of growth at higher temperatures on mineral sulphides than any other known isolates.

1.2 - Archaea and phylogeny

Until the 1970s, life was considered to be segregated into two kingdoms, the eukaryotes and the prokaryotes. However, since that time a third major division of life has been identified, which has radically altered views of the classification and evolution of all organisms (Woese and Fox, 1977). This relatively new domain is the Archaea, comprising organisms previously known as the archaebacteria, which are remarkable for their tolerance to what are considered extreme conditions of life. It is thought that the divergence of the archaea as a discrete group from the eukaryotes may have taken place after the divergence of the eubacteria from a group containing both archaea and eukaryotes (Klenk and Doolittle, 1994). It has been suggested that the low rate of sequence change in archaea and their relationship to eukaryotes may allow studies of archaea to shed light on the makeup of primordial eukaryotes (Marsh *et al.*, 1994).

The archaea may be roughly split into three major groups on the basis of their environmental niches. These groups are the halophiles, the methanogens and the thermophiles. The thermophiles include those organisms which are the subject of this thesis, *Sulfolobus* strain LM and the *Sulfolobus*-like strain HT. More rigorous studies of the phylogeny of the archaea have compared their whole cell protein profiles, molar % GC contents, DNA:DNA hybridisation characteristics and the sequences of 16S rRNA.

The traditional classification of organisms by comparison of their morphologies and growth characteristics was found not to be reliable for prokaryotes, and provided little phylogenetic information. In contrast, the comparison of macromolecules that are functionally homologous in all known organisms has allowed the establishment of a universal phylogenetic framework. The macromolecule most extensively investigated in this respect is the small subunit (16S) rRNA. This exhibits the necessary characteristics of a useful phylogenetic marker in that it is ubiquitous, of essential function and

contains sections ranging between highly variable and invariant, thus allowing it to be used for classification of both highly related and very dissimilar organisms. rRNA classification systems have been found to be in fairly good agreement with other phylogenetic markers such as elongation factors, ATPase subunits and RNA polymerases. It also appears not to have been the subject of lateral gene transfer, probably due to its essential function (Ludwig and Schleifer, 1994).

16S rRNA classification of the archaea has resulted in the recognition of two main groups (Olsen et al.,1985). The euryarchaeota include the methanogens, the halobacteria and the wall-less thermophile, *Thermoplasma acidophilum*. The crenarcheota includes the majority of the thermophiles, as well as some planktonic isolates. The evolutionary separation of these two archaeal groups exceeds that of any divisions yet characterised within the eukaryotic or eubacterial urkingdoms (Fox *et al.*, 1980). This implies a diversity of characteristics amongst the archaea at least as great as amongst the eubacteria.

Of the thermoacidophilic archaea, the genus which has received most attention is *Sulfolobus*. There are, at present, only two published 16S rRNA sequences for species of this genus. These are for *Sulfolobus acidocaldarius* (Olsen *et al.*, 1985) and *Sulfolobus shibatae* (Grogan *et al.*, 1990). The *Sulfolobus acidocaldarius* sequence was originally submitted as *Sulfolobus solfataricus*, highlighting a confusion in strain identities. Certain culture collection *Sulfolobus* strains initially characterised as sulphur oxidizers now appear unable to oxidize sulphur, adding to the confusion which a more comprehensive database would help to resolve. The strains which have been most studied, and for which 16S rRNA sequences are available, do not oxidize iron or mineral sulphides. The strains which do oxidize iron, and are therefore of potential industrial use in mineral leaching, are relatively little studied and have not all been characterised sufficiently for reliable identification or classification.

The potentially useful strains for mineral processing include the HT isolate, a *Sulfolobus*-like thermoacidophile isolated from an Icelandic hot spring. This strain has shown a high optimum temperature for mineral leaching of 85°C, thus making it potentially the most useful of the mineral sulphide oxidising archaea (Norris and Owen, 1992). The LM isolate is another *Sulfolobus*-like thermoacidophile also isolated from an Icelandic hot spring which appears to be of the same species as an isolate designated strain BC, obtained from an acidic coal spoil heap in the United Kingdom. The LM and BC strains have been the subject of much research on mineral leaching, and the LM strain is the only *Sulfolobus* isolate whose autotrophic growth has been studied in any depth (Norris *et al.*, 1986; Wood *et al.*, 1987; Norris *et al.*, 1989; Nixon and Norris, 1992). It may (Norris, pers. comm.) also be of the same species as the more recently isolated strain named *Sulfolobus metallicus* (Huber and Stetter, 1991).

1.3 - Chemolithotrophy in archaea; the assimilation of CO_2

The specific areas of metabolism enabling the mineral-oxidising archaea to grow chemolithotrophically are sulphur and reduced sulphur compound oxidation, ferrous iron oxidation and CO_2 fixation. Elucidation of the mechanisms of and components involved in chemolithotrophic growth by a molecular biological approach has certain advantages with potentially commercially-useful organisms. It requires the identification and characterisation of key proteins and their genes which might in future become targets for manipulation to achieve strain improvement. One problem avoided by such a course of investigation is that of having to obtain large quantities of purified proteins for direct characterisation from low biomass-yielding chemolithoautotrophs, especially where growth upon an energetically poor substrate such as ferrous iron is involved. The investigation of the iron oxidation system from the mesophilic eubacterium *Thiobacillus ferrooxidans* has illustrated this point. The characterisation of the key iron oxidase enzyme was facilitated by N-terminal protein microsequencing

followed by the cloning and sequencing of its gene and subsequent computer analysis (Kusano et al., 1993). Genetic manipulation of the moderately thermophilic iron oxidizing eubacteria and the archaeal thermoacidophiles cannot even be considered at present due to the lack of reliable DNA transformation systems for these organisms.

Two genes encoding proteins directly related to chemolithoautotrophic growth mechanisms in thermoacidophilic archaea have so far been sequenced. These are the sulphur oxygenase / reductase from *Desulfurolobus ambivalens* (Kleitzin, 1992) and the aa_3 cytochrome oxidase from heterotrophically grown *Sulfolobus acidocaldarius* (Lübben *et al.*, 1992). The possibility exists that this terminal oxidase is active during autotrophic iron oxidation and sulphur compound oxidation.

The two major areas of chemolithotrophic metabolism noted earlier which have not been the subject of molecular biological investigation are CO_2 fixation and iron oxidation.

The pathway of carbon dioxide fixation in *Sulfolobus* has not been resolved in detail. The enzymes of the Calvin cycle are known to be absent from *Sulfolobus* and similar mineral-oxidising archaea (Brierley and Brierley, 1986). They efficiently oxidize mineral sulphides when supplied with air levels of CO_2 , 0.033 % (v/v), although the rate of pyrite oxidation is reduced by 20 % in comparison with cells provided with 1 % (v/v) CO_2 (Norris *et al.*, 1989). The concentration of CO_2 necessary to maintain optimal levels of mineral dissolution from mineral sulphides in bioreactors may influence the cost effectiveness of the application of microorganisms.

There are at present three main recognised pathways of CO_2 fixation during autotrophic growth in prokaryotes. The best known is the reductive pentose phosphate cycle or Calvin cycle. This was assumed to operate in all autotrophic plants and

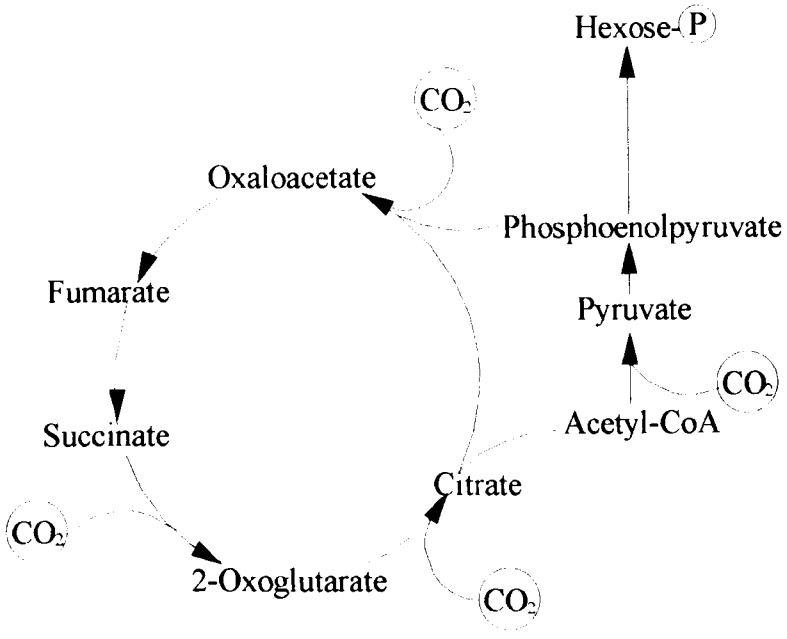
microorganisms. However, it is now clear that there are at least two additional systems of CO_2 fixation operating in archaea and eubacteria.

The Calvin cycle is the mechanism by which the eubacterial mineral oxidising bacteria fix CO_2, both the mesophilic Gram negative *Thiobacillus ferrooxidans* and the moderately thermophilic Gram positive bacteria such as *Sulfobacillus* species (Wood and Kelly, 1985). This mechanism is confined to the aerobic and facultatively aerobic eubacteria and to chloroplasts.

The second mechanism of CO_2 fixation appears to be confined to the strictly anaerobic and microaerophilic eubacteria and the methanogenic archaea. This is the reductive acetyl-CoA pathway. In this pathway, acetyl-CoA is produced from two molecules of CO_2. The carboxyl carbon of the acetyl-CoA formed is derived from carbon monoxide, formed by the reduction of CO_2, or by the direct binding of carbon monoxide to acetyl-CoA synthase (Fuchs, 1986). This is therefore an inappropriate mechanism for CO_2 fixation during aerobic growth, as exhibited by *Sulfolobus*.

The third mechanism of CO_2 fixation is the one likely to be most relevant to the thermoacidophilic archaea. This is the reductive citric acid cycle (see Fig. 1.1 overleaf). It has been identified in green sulphur eubacteria such as *Chlorobium limicola*, thermophilic hydrogen oxidising eubacteria such as *Hydrogenobacter thermophilus*, sulphate reducing eubacteria such as *Desulfobacter hydrogenophilus* (Fuchs, 1988) and also in the archaeon *Thermoproteus neutrophilus* (Schafer et al., 1989).

Figure 1.1 The reductive citric acid cycle for autotrophic CO_2 fixation in *Thermoproteus neutrophilus* (after Danson, 1989)



Where it has been characterised, this cycle has been shown to use reverse reactions of the oxidative carboxylic acid cycle. Key enzymes of this pathway are 2-oxoglutarate synthase, which differs from 2-oxoglutarate dehydrogenase in composition, electron donor and other catalytic properties, and ATP citrate lyase, responsible for the cleavage of citrate into acetyl-CoA, the product of the cycle, and oxaloacetate, to restart the cycle. A modified reverse citric acid cycle would be expected to be reversible, and utilised for CO_2 fixation under autotrophic conditions and oxidation under heterotrophic conditions.

The assimilation of the acetyl-CoA produced by the reductive carboxylic acid cycle and reductive acetyl-CoA pathway has been shown to occur by the action of a pyruvate synthase, which reductively carboxylates acetyl-CoA to pyruvate. The pyruvate then proceeds either to oxaloacetate by pyruvate carboxylase or to

phosphoenolpyruvate (PEP). PEP may be then used for gluconeogenesis, or for anaplerotic CO_2 fixation via PEP carboxylase or PEP carboxykinase (Fuchs, 1988).

It has been suggested that the reductive conversion of oxaloacetate into succinate via malate and fumarate may have emerged in association with hexose metabolism to regenerate NAD+. The formation of acetyl-CoA from pyruvate via an oxidoreductase may have evolved in early anaerobic cells, leading to the evolution of the other half of the cycle. The two halves of the cycle would be linked by 2-oxoglutarate synthase, leading to the formation of a complete cycle. The reductive carboxylic cycle, not the oxidative cycle, may represent the primitive pathway in evolutionary terms (Danson and Hough, 1991).

The activity of this pathway may be indicated by the early labelling of alanine in $^{14}CO_2$ assimilation experiments, showing the formation of pyruvate from acetyl-CoA. This early labelling of alanine was not found in experiments using the thermoacidophilic archaea *Acidianus brierleyi* (Kandler and Stetter, 1981) and *Sulfolobus* LM (Norris *et al.*, 1989). However, the action of an as yet undefined form of the reductive carboxylic acid cycle was indicated. The studies using *Sulfolobus* highlighted carbon dioxide fixation activity by an ATP dependent acetyl-CoA carboxylase and a PEP carboxylase. Therefore, the nature of this modified reverse carboxylic acid cycle in mineral-oxidising archaea remains to be characterised. Favouring a molecular approach as previously described, the putative ATP-dependent acetyl-CoA carboxylase gene was to be cloned and sequenced in an attempt to further understand the nature of this modified CO_2 fixation mechanism in *Sulfolobus*.

1.4 - Iron oxidation by acidophilic microorganisms

The mechanism of ferrous iron oxidation in the archaea is unknown. A general picture of the types of components likely to be involved in such a process may be gathered from the studies of this process in eubacteria.

The study of iron oxidation in acidophiles has most often been in the context of their application in mineral sulphide oxidation for metal extraction, i.e. bioleaching.

At present the most commonly used bioleaching bacterium is a mesophile, *Thiobacillus ferrooxidans*. This bacterium was discovered in acid drainage from a pyritic coal mine (Colmer and Hinkle, 1950). It possesses some remarkable properties. It derives energy from the oxidation of a range of reduced sulphur compounds including sulphides, elemental sulphur, thiosulphate, and sulphite, producing sulphate as an end product. It grows autotrophically, oxidising ferrous iron to ferric iron. It is also an acidophile, the preferred pH range being from pH 1.5 to pH 2, decreasing as the bacterium produces sulphuric acid.

As work has proceeded with the exploitation of *Thiobacillus ferrooxidans*, a wider range of bacteria has been discovered sharing its sulphur and iron oxidation properties and acid tolerance (Norris, 1990). At present three major groupings of these organisms have been recognised.

The Gram negative mesophiles include *Thiobacillus ferrooxidans*, and *Leptospirillum ferrooxidans*. These organisms have growth temperature optima between 30°C and 37°C.

The Gram positive moderate thermophiles have been studied to a lesser extent than the Gram negative mesophiles. These include *Sulfobacillus thermosulfidooxidans*

(Karavaiko *et al.*, 1990) and strains such as BC1, TH3, ALV and N. Their temperature optima range between 45°C and 50°C (Norris and Barr, 1985).

The archaeal extreme thermoacidophiles which oxidize iron include *Sulfolobus* LM, the *Sulfolobus*-like HT strain, *Metallosphaera sedula* and *Acidianus brierleyi*. They have optimal growth temperatures between 60°C and 85°C (Stetter, 1986).

1.5 - The application of microbial iron oxidation in mineral leaching

The industrial application of bacterial involvement in the solublisation of metals from mineral sulphides (Hutchins *et al.*, 1986; Brierley and Brierley, 1986; Kelly, 1988) has proceeded for many years as low technology heap and dump leaching operations for uranium and copper extraction. More recently, however, bioreactors have been introduced for the leaching of precious metals from ores (Hutchins *et al.*, 1987; Spencer *et al.*, 1989). Bioreactors offer advantages over earlier methods due to the degree of control that can be exerted over the conditions affecting bacterial growth and activity. Bioleaching avoids the SO_2 emissions produced by the usual smelting processes, and has allowed metal extraction from low grade and refractory ores.

There are currently a number of bioreactors currently in commercial use or undergoing construction for the treatment of gold bearing mineral sulphides. These include the South African Fairview mine (processing 35 tons of ore per day), the Brazilian Sao Bento mine (150 tons / day), two Australian mines (40 and 115 tons / day) and the Ashanti Goldfields Corporation plant in Ghana (720 tons / day).

The reactions involved in mineral bioleaching are complex. These include the direct bacterial attack on the sulphide moiety of the ore, resulting in sulphuric acid production and the acidification of the leaching environment. Ferric iron, which can

remain in solution at low pH, can then chemically interact with the ore, resulting in the solublisation of its metallic content. The ferric iron is recycled via the bacterial oxidation of ferrous iron (Hutchins *et al.*, 1986).

One effect slowing the bacterial solublisation of ores is the deposition of precipitate layers on the mineral surfaces. This effect appears to be greatly diminished at higher temperatures, at which the rate of mineral dissolution has been shown to be greatly increased (Norris and Parrot, 1986). As the bacterial leaching of ores is an exothermic reaction, reactors using mesophilic bacteria require expensive cooling. High temperature leaching reduces the extent to which this is necessary. The use of high temperatures obviously requires the use of a thermophilic bacterium.

Additional roles for these bacteria have been envisaged in the controlled desulphurization of coal and decontamination of industrial slag heaps. Also, they may have potential for biotransformations, the use of acid stable enzymes such as proteases and amylases and the use of thermostable enzymes (Norris, 1992).

Due to their phylogenetic diversity, various physiological and biochemical features of these groups of bacteria, such as their mechanisms of iron oxidation and CO_2 fixation appear to be dissimilar (Norris, 1990). Factors limiting the application of different species include their optimum growth temperatures, CO_2 requirement, metal sensitivity, ferric iron inhibition and sensitivity to agitation in reactors containing high mineral densities.

1.6 - Mechanisms of iron oxidation in acidophiles

The system of ferrous iron oxidation in the Gram negative *Thiobacillus ferrooxidans* has been the most studied and various protein components of the pathway for the transport of electrons from ferrous iron have been isolated (Ingledew, 1982; Tikhonova *et al.*, 1987). Iron oxidation presumably occurs in the periplasm, since at pH 3, oxidation produces soluble ferric iron, whereas at pH 5.5, inside the cytoplasm, ferric iron would precipitate. Initially it was thought that ferrous iron was oxidized directly by rusticyanin (Cox and Boxer, 1978), an abundant blue copper protein found in the periplasm. Due to the relatively slow kinetics of electron transfer from ferrous ions in solution to rusticyanin, it has been suggested that bound iron atoms are important as an intermediate. More recently, however, Fe(II) cytochrome c oxidoreductase has been isolated (Fukumori *et al.*, 1988) and its gene cloned and sequenced (Kusano *et al.*, 1992). This appears to directly oxidize ferrous ions, electrons subsequently being transferred to cytochrome c-552 (Sato *et al.*, 1989) and rusticyanin. From this stage, electrons are transferred to an a-type cytochrome, embedded in the cytoplasmic membrane, and acting as the cytochrome oxidase (Yamanaka *et al.*, 1989), enabling reduction of oxygen inside the cytoplasm (Poole *et al.*, 1988). The exact relationships between these proteins have yet to be completely elucidated.

Iron oxidation in the thermophiles, by contrast, has not been studied to a similar extent. The mechanism of iron oxidation used by *Thiobacillus ferrooxidans* is by no means ubiquitous. *Leptospirillum ferrooxidans*, another Gram negative mesophile, has been analysed spectrophotometrically, and does not appear to possess a similar terminal electron acceptor to that of *Thiobacillus ferrooxidans*. In addition, rusticyanin is not found in *Leptospirillum ferrooxidans*, but a novel acid stable cytochrome has been isolated (Hart *et al.*, 1991).

Moderately thermophilic iron-oxidising bacteria, being Gram positive, have a radically different cell wall architecture from *Thiobacillus ferrooxidans* and *Leptospirillum ferrooxidans*, possessing no enclosed periplasm. Therefore, initial ferrous iron-oxidising proteins are expected to be different (Ferguson, 1988). This is confirmed by the absence of large amounts of acid-stable electron carriers analagous to rusticyanin and the *Leptospirillum ferrooxidans* cytochrome (Barr *et al.*, 1990).

The archaea also possess a different cell wall structure, and also a radically different cell membrane structure, based on bipolar lipids so would be postulated to possess another mechanism of iron oxidation.

Archaea share the basic characteristics of energy conversion by electron transport phosphorylation with eubacteria and eukaryotic organelles, coupling intracellular proton flux to ATP synthesis. Under aerobic conditions respiratory enzymes act as redox energy driven transmembrane proton pumps. Thermoacidophilic archaea require effective mechanisms for generation and maintenance of proton gradients, *Sulfolobus* possesses an almost neutral cytoplasmic pH, but grows at pH 1.5. During heterotrophic aerobic respiration *Sulfolobus* forms a proton motive force of -150 mV, sufficient for supporting steady state ATP synthesis.

The respiratory chain proteins of archaea have not all been studied on a molecular level even in heterotrophically-grown archaea. The only possible respiratory chain protein identified as being unique to iron oxidation is a membrane-associated acid-stable novel cytochrome. This has been identified in *Sulfolobus* strains HT and LM, *Metallosphaera sedula* and *Acidianus brierleyi* (Barr *et al.*, 1990).

Two respiratory terminal oxidase complexes of heterotrophically grown *Sulfolobus acidocaldarius* have been isolated, partially biochemically purified, and their gene sequences reported. The first complex to be identified was the Sox ABCD

complex, showing redox activity at 603 nm (Anemüller and Schäfer, 1989; Wakagi *et al.*, 1989). This contains two cytochrome oxidase subunits, a *b* type cytochrome and two peptides of unknown function. These proteins show homology to subunits I and II of mitochondrial cytochrome *c* oxidase and to cytochrome *b*, a subunit of mitochondrial and bacterial cytochrome *c* reductases and the chloroplast *b6f* complex. The haems of the Sox ABCD complex are of a novel variety, possessing geranyl-geranyl as the long hydrophobic side chain instead of farnesyl (Lubben *et al.*, 1994).

The Sox M complex has recently been investigated, showing redox activity at 562 nm. This complex is similar to that of Sox ABCD in that it appears to be a hybrid containing components of the cytochrome *bc* and cytochrome oxidase complexes. It also contains a Rieske iron-sulphur protein and a blue copper protein.

The subunit compositions of the respiratory complexes in archaea therefore appear quite different from those of eubacteria. This has some implications for the evolution of aerobic metabolism. These rest on the hypothesis that every protein homologous in archaea and eubacteria must have been present in their last common ancestor, unless lateral gene transfer has occurred between these two groups during evolution (Benner *et al.*, 1989). Lateral gene transfer may be detected by careful analysis of phylogenetic trees produced by comparison of related proteins from different organisms. It is thought that all key proteins of energy transduction in eukaryotes are derived from the eubacteria via mitochondria and chloroplasts. The last common ancestor, therefore, possessed a mitochondrial type oxidase which received electrons from cytochrome *c* or a blue copper type or related metalloprotein. As the photosystems are found only in eubacteria, no chlorophyll based photosynthesis having been found in the archaea, this implies that aerobic metabolism evolved earlier than photosynthesis (Castreana *et al.*, 1995).

In *Sulfolobus* LM, respiratory iron oxidation may perhaps be linked to the Sox ABCD terminal oxidase. However, the nature of the iron oxidases and electron carrier proteins, including the novel cytochrome, and their possible relationship to a terminal oxidase complex remains to be elucidated.

The nutritional versatility of the thermophiles may allow the identification of iron oxidation components by growth on different substrates and thus examination of the bacteria when grown with or without iron.

The Gram positive moderate thermophiles, strains BC1, LM2, TH3 and ALV have been investigated spectrophotometrically, all showing major absorbance peaks at approximately 440, 563, and 603 nm (Barr *et al.*, 1990). This pattern varied little in the presence or absence of iron oxidation, and, as expected, no obvious small soluble electron carriers, corresponding to rusticyanin, were found. Optical spectrophotometry has also been carried out on some archaea, such as *Sulfolobus* LM and *Metallosphaera sedula* (Barr *et al.*, 1990), showing absorbance peaks at about 483, 535, 572 and 603 nm. The 603 nm peak has been shown to be the terminal oxidase aa_3-type cytochrome, which has been isolated and its gene cloned and sequenced (Anemüller and Schäfer, 1990; Lübben *et al.*, 1992), and is not unique to iron oxidation. However, the 572 nm peak appears to be closely related to the functioning of iron oxidation. This may represent a novel cytochrome involved in iron oxidation.

An additional aspect important during iron oxidation and mineral leaching applications is that of resistance to potentially toxic metals (Brierley and Brierley, 1986; Norris and Owen, 1992). Specific metal binding molecules, such as metallothioneins, may be produced in response to increased concentrations of metal ions. Preliminary studies with *Sulfolobus solfataricus* have shown zinc and copper induction of certain proteins (Scudiero *et al.*, 1992). However, no further details of these metal induced proteins have emerged. Some species of the thermoacidophilic

archaea have been shown to be sufficiently sensitive to copper to preclude their use in commercial bioleaching, whereas other strains have been adapted to tolerate very high levels of copper. For example *Metallosphaera sedula* is killed by copper at 5 g / l (w/v), but *Sulfolobus* LM tolerates 45 g / l (w/v) (Norris and Owen, 1992). This illustrates the importance of metal resistance to application of these isolates.

1.7 - Summary of project aims

1 - The cloning and sequencing of the gene encoding the 16S rRNA from *Sulfolobus*-like strain HT, and subsequently its phylogenetic analysis.

2 - The characterisation of a protein highly expressed during CO_2-limited autotrophic growth of *Sulfolobus* LM.

3 - The characterisation of components over-expressed during the autotrophic growth of *Sulfolobus* LM on ferrous iron.

4 - The isolation and further characterisation of a novel cytochrome over-expressed during the autotrophic growth of *Sulfolobus* LM on ferrous iron.

CHAPTER 2

METHODS AND MATERIALS

2.1 - Chemicals

All chemicals apart from those noted were obtained from the following companies ;-

Sigma Chemical Co. Ltd., Poole, Dorset, U.K.
Fisons Scientific Equipment Ltd., Loughborough, U.K.
British Drug Houses (BDH) Ltd., Poole, Dorset, U.K.

The following purified waters were used ;-

Double distilled water - used for general applications

Elgastat deionized water via Spectrum cartridge types SC 32 and SC 1 (Elga Ltd., High Wycombe, Buckinghamshire, U.K.) - used for media production

Super-O-water at resistivity of >10 MΩ via a deionizer (Purite Ltd., Thame, Oxfordshire, U.K.) and four further cartridges (Millipore Co., Bedford, Massachussets, U.S.A.) being one Super-C organic adsorption cartridge, two Ion-Ex deionization cartridges and one Millitube MF cartridge - used for molecular biological applications.

2.2 - Strains used

Two different strains of *Sulfolobus* were used, *Sulfolobus* HT for the 16S rRNA work , and *Sulfolobus* LM (see section 1.2). Both were available from stocks maintained at Warwick. These had been grown using sulphur and pyrite media (see section 2.3), stored at room temperature and subcultured monthly.

2.3 - Growth conditions for *Sulfolobus* LM

100 ml stock cultures were incubated in 250 ml conical flasks shaken at 100 rev/min in a Gallenkamp orbital shaker at 65°C. For autotrophic growth, or to scale-up to 20 l fermenters, 1 l cultures were grown in 2.5 l flasks. The following mineral salts medium was used.

Mineral salts medium	g/l
Magnesium sulphate heptahydrate	0.4
Ammonium sulphate	0.2
Potassium chloride	0.1
Di-potassium hydrogen sulphate	0.1

Added to the mineral salts as required were the differential components;

For growth on ferrous iron;-

Ferrous sulphate heptahydrate	13.9 g/l
Potassium tetrathionate	0.015 g/l

For growth on tetrathionate;-

Potassium tetrathionate	0.3 g/l
Ferrous sulphate heptahydrate	0.0139 g/l

For growth on elemental sulphur;-

Flowers of sulphur	5 g/l
Ferrous sulphate heptahydrate	0.0139 g/l

For growth on arsenopyrite;-

Arsenopyrite As-Ls 19 (Olympia concentrate, Greece)	5 g/l

Cultures were gassed with air or with elevated levels of carbon dioxide, achieved by maintaining a constant supply of filtered 1% (v/v) CO_2 in air, which was continuously bubbled through the medium. This was supplied by diluting 5% (v/v) CO_2 in air (Distillers MG Ltd., Reigate, Surrey, U.K.) with compressed air using a flowmeter (Platon Ltd., Basingstoke, Hampshire, U.K.).

Occasionally cultures were gassed with elevated CO_2 levels by 5% (v/v) CO_2 in air when an excess of CO_2 was required.

pH values were set, using a PHM 62 pH meter (Radiometer, Copenhagen, Denmark) and an ATI pH probe (Russell, Fife, U.K.) with 5% (v/v) sulphuric acid, to pH 1.5 for ferrous iron medium, pH 3.0 for sulphur medium, and pH 2.0 for tetrathionate medium and for pyrite medium.

Mineral salts media were autoclaved at 121 ^{o}C for 15 min in a pressure cooker or autoclave. Ferrous iron and tetrathionate solutions were autoclaved at 115 ^{o}C for 10 min, then added to the media. Flowers of sulphur were added directly to the mineral salts media, Tyndalised by autoclaving at 109 ^{o}C for 5 min once on each of three successive days. All media was brought to the correct temperature for the growth of cultures before inoculation.

Inocula of 5% (v/v) were used, and cultures grown until mid to late exponential phase. Growth was determined either by measuring absorbance at 440 nm using an SP6-200 spectrophotometer (Pye-Unicam Ltd., Cambridge, U.K.), or determination of percentage iron oxidation by titration of residual ferrous iron with ceric sulphate using 1:10 phenanthroline-ferrous sulphate complex as an indicator. Organisms were examined by phase contrast microscopy.

2.4 - Large scale growth of *Sulfolobus* LM in 20 l vessels

Where biomass was required for protein purification, 20 l cultures were grown. Media and temperature conditions were as detailed above. The apparatus used consisted of 20 l flat bottomed vessels sealed with a bung pierced by a sparger, a glass tube for a feedback thermometer (Russell), and a port for connection to a Leibig condenser. Where required, the media was sparged with filtered 1% CO_2 in air at a flow rate of 10 l per minute, otherwise with filtered air at the same flow rate. Due to the high temperature and flow rate, the rate of evaporation was also high. This was remedied by use of two large Leibig condensers. The condenser port could also be used for taking samples. Agitation, and consistent high temperature were achieved by the use of a Combimag RET heater stirrer (IKA) connected to a feedback thermometer. Good contact between the bottom of the vessel and a supporting metal sheet between it and the heater stirrer hot plate was essential.

After autoclaving, all vessels were allowed to cool to the correct temperature before inoculation with 5 - 10% of the final volume. Samples were monitored for pH , optical density and degree of oxidation until mid to late log phase, then cells were harvested. A cross flow filter (Flowgen) connected to a peristaltic pump was used to concentrate the cells into a volume of less than 1.5 l, which was then centrifuged for 10 min at 4420 g using a JA 10 rotor in a J2-21 M/E Beckmann centrifuge. Cells were resuspended in pH 2 (H_2SO_4) water, then spun in test tubes in a bench top centrifuge to remove precipitated material such as iron hydroxides or sulphur particles. Cells were stored at -20^{o}C.

2.5 - Cell lysis

Cell pellets of *Sulfolobus* LM were lysed for membrane fraction preparations and genomic DNA preparations. Various methods of lysis were attempted. These were; raising the pH of the lysis buffer from 2 to 8, sonication for 3 x 30 sec at 6 microns at 0°C, and using a French press (American Instrument Co., Silverspring, Maryland, U.S.A.) at 1000 Pa. Results were monitored by microscopic examination under phase contrast, lysis of *Sulfolobus* being readily apparent due to the loss of their coccoid morphology. Neither method was found to be entirely satisfactory alone, but combining the methods gave an effective procedure.

The cell pellet was first resuspended in water, using a hand held glass homogeniser, washed by being microfuged and resuspended twice, then resuspended in 50mM Tris /1mM EDTA at pH 8. As the pH was increased, cells began to lyse, tending to lower the pH again. Therefore, dilute sodium hydroxide was used to correct the pH during the procedure, care being taken not to allow the pH above 8. The partially lysed mixture was then French pressed six times, during which time samples were monitored by microscopy and the pH corrected as necessary.

Where relatively large quantities of cells were to be lysed, a sonication step was found to be advantageous, and was carried out before suspension in pH 8 buffer. This was necessary to completely homogenise the mix of resuspended cells. Throughout the lysis procedure, all solutions were kept on ice.

2.6 - Preparation of genomic DNA from *Sulfolobus* species

For the preparation of genomic DNA from both *Sulfolobus* LM and HT, the following procedure was used. Cells were lysed as above, with addition of NaCl to 400 mM and sucrose to 750 mM to the lysis solution. SDS was added to 1% (w/v) final concentration, and proteinase K (Gibco-BRL) to 100 μg/ml. The solution was incubated at 55°C with gentle shaking for 1 hour.

The lysate was diluted with 2 volumes of EDTA at 250mM, then the protein extracted. Protein extraction was carried out by addition of 1 volume of phenol (equilibrated in STE) , inverting the container very gently, so as to avoid excessive shearing of the DNA strands, then centrifuging in a Beckman JA 20 rotor at 39200 g for 30 min. The lower, phenol containing, layer was discarded, and extraction repeated.

The DNA solution, made up to 10 ml with sterile TE (10 mM tris, 1mM EDTA) was loaded into a polypropylene tube containing 30 g of CsCl dissolved in 20 ml TE and 1 ml of ethidium bromide solution (10 mg/ml) This was ultracentrifuged using a Beckman L8-60M ultracentrifuge with a Vti 50 rotor at 120000 g for 16 h at 20°C. The remaining protein was precipitated, while nucleic acids formed two discrete bands, the lower being RNA, the upper DNA, these were visualised under ultraviolet illumination. The DNA was removed by being allowed to drip through a wide bore needle very slowly.

To remove ethidium bromide, the DNA solution was mixed with an equal volume of butanol equilibrated with TE. The mixture was inverted gently and the contaminated layer was removed. This was repeated until no further pink colour was seen, then the DNA solution was dialysed 3 times against 1 l TE at 4°C for 2-12 h. The dialysis step removed butanol which might inhibit later enzyme reactions. The

DNA solution was then concentrated using PEG 2000 to the required volume. Samples were measured for absorbance at 260 nm and 280 nm using a DU 70 spectrophotometer (Beckman) , A_{260} divided by 50 approximating to DNA concentration (μg/ml) , and A_{260} / A_{280} of near 1.7 indicating purity of the DNA. The solution was stored at 4°C.

2.7 - Restriction enzymes and minigels

Various restriction endonucleases were used. These were obtained from Gibco-BRL (Gaithersburg, Maryland, U.S.A.) or Amersham International (Amersham, Buckinghamshire, U.K.). Usually used were ;-

1-2 μg of DNA
1-5 units of enzyme(s)
10 μg RNAse (Gibco-BRL) where required
2 μl of relevant 10 x restriction buffer (Gibco-BRL)
made up to 20 μl with sterile distilled water
incubated at recommended temperature for 1-2 h

Gel electrophoresis was carried out on restricted DNA fragments using minigel apparatus (Flowgen Ltd., Sittingbourne, Kent, U.K.). DNA was loaded in a 5% (v/v) glycerol / 0.08% (w/v) bromophenol blue / 0.08% (w/v) xylene cyanol mix. The constituents of the gels were;-

0.8% to 1.2% (w/v) agarose (Sigma)
1 x TBE buffer (Sambrook *et al.*, 1989)
Ethidium bromide at 200 mg/ml for visualisation of DNA

Gels were run in 1 x TBE, bands of DNA were visualised under UV illumination (Ultra Violet Products Inc., San Gabriel, California, U.S.A.) and photographed on Polaroid 665 film with a CU5 camera (Polaroid, St. Albans, Hertfordshire, U.K.).

Molecular weight markers (Gibco-BRL) of Lambda phage restricted with *Hin*d III or *Pst* I were used. The *Hin*d III digested markers gave bands at 23.1, 9.4, 6.6, 4.4, 2.3, 2.0, 0.56 and 0.12 kb. The *Pst* I digested markers gave bands at 11.5, 5.0, 4.7, 4.5, 2.8, 2.6, 2.5, 2.4, 2.1, 2.0, 1.7, 1.2, 1.1, 0.8, 0.5, 0.47, 0.45, 0.34, 0.26, 0.25, 0.22, 0.21, 0.2, 0.16 and 0.15 kb.

2.8 - Maxigels and Southern blots

Genomic DNA (10µg) was digested with a range of restriction enzymes, with incubation being allowed to continue for at least 2 hours at 37 ^{o}C. A maxigel was poured consisting of 350 ml of 0.8% (w/v) agarose gel as above. This was run for 24 hours at 20 mA in a gel tank containing 2.5 l of TBE. The gel was photographed, then treated to promote efficient DNA transfer. This treatment consisted of ;-

- gel soaked for 10 min in 0.2M HCl, then rinsed in deionised water.
- DNA denatured by soaking gel for 45 min in 1.5 M NaCl / 0.5 M NaOH
- Gel rinsed with deionised water, then neutralised in 1.5 M Tris/HCl (pH 7.4) / 1.5 M NaCl for 30 min, then solution changed for another 15 min incubation.

The gel was then set up for Southern blotting using 10x SSC (Sambrook *et al.*, 1989) as the transfer buffer and a pre-wetted positively charged nylon membrane (Boehringer Mannheim, Mannheim, Germany) to bind the DNA. After allowing transfer to proceed for 12-24 h, the DNA was fixed to the nylon membrane while still

damp using UV crosslinking at 254 nm from a U.V.Stratalinker 2400 (Stratagene Ltd., Cambridge, U.K.). The membrane was then air dried before further experimentation.

2.9 - Non-radioactive oligonucleotide labelling

Consensus oligonucleotides designed using known sequence were used to probe Southern blots and genomic libraries. These oligonucleotides (16-21-mers) were labelled non-radioactively using the DIG oligonucleotide 3' end labelling kit (Boehringer) according to the manufacturer's protocol. This kit utilises the enzyme terminal transferase to add a single digoxigenin labelled dideoxy-UTP nucleotide to the 3' end of each oligonucleotide. Efficiency of labelling was checked by performing a dilution series on an aliquot of the labelled probe, then spotting each of the dilutions onto nylon membrane versus a control then performing the detection procedure (see section 2.12).

2.10 - Non-radioactive labelling of DNA

Double stranded DNA probes longer than 100 base pairs were labelled using the DIG DNA labelling kit (Boehringer), according to the manufacturer's protocol. This kit uses a standard random priming reaction to incorporate digoxigenin labelled dUTP into complementary strands of the denatured template DNA using DNA polymerase I (Klenow fragment).

2.11 - Hybridisation

Hybridisation is the reaction by which the labelled probe DNA binds to the target DNA which is itself immobilised on a filter such as nylon membrane. The first step of hybridisation was the prehybridisation. The filter was incubated at 65°C in hybridisation solution containing blocking reagent (Boehringer) which minimises non-specific binding of the probe to the filter by itself blocking non-specific nucleic acid binding sites. Hybridisation solution consisted of ;-

5x SSC

0.1% (w/v) N-lauroylsarcosine

0.02% (w/v) SDS

1 % blocking reagent (Boehringer)

Prior to adding to the hybridisation mix, double stranded probes were denatured by heating to 95°C. The hybridisation reaction was usually carried out at 65°C for long DNA probes, to obtain maximum stringency. Oligonucleotide probes were usually hybridised at a temperature 10°C lower than their theoretical melting temperature, which could be approximated by allowing 2°C for each A-T bond and 3°C for each G-C bond. In practice, however, test hybridisations were carried out at a range of temperatures from 37°C to 60°C to find the optimum in terms of sensitivity and specificity. Hybridisations were carried out in sealed plastic bags in an agitated temperature controlled waterbath or preferably in sealed rotating tubes in a Mini 1 oven (Hybaid Ltd., Teddington, Middlesex, U.K.). Hybridisation was allowed to occur for 12-18 h, then unbound probe was washed off the filters at the hybridisation temperature. Washing solution was made up of 0.1% SDS plus SSC in concentrations varying from 0.1x to 5x, which altered the stringency. As for the hybridisation temperature, the stringency of washing required optimisation in each case.

2.12 - Detection of digoxigenin labelled DNA

Two different methods of detecting digoxigenin labelled DNA were used, colorimetric and luminescent detection. Colorimetric detection was provided by the DIG nucleic acid detection kit (Boehringer). Luminescent detection was provided by the DIG luminescent detection kit (Boehringer). Both were used according to manufacturer's protocols.

Both of these kits utilise polyclonal anti-digoxigenin antibody Fab fragments conjugated to alkaline phosphatase. The antibody-enzyme conjugate binds specifically to the digoxigenin labelled probe. In the colorimetric reaction the alkaline phosphatase catalyses NBT solution (nitroblue tetrazolium salt) and X-phosphate solution (5-bromo-4-chloro-3-indolyl phosphate toluidinium salt) to produce a purple / blue coloured precipitate, which can be directly visualised. In the luminescent detection method, the conjugated alkaline phosphatase catalyses the release of luminescence from the reagent lumigen PPD [4-methoxy-4-(3-phosphate-phenyl)-spiro(1,2-dioxetane-3,2'-adamantane) disodium salt]. This luminescence was detected by exposure to Fuji RX photographic film (Fuji Film Co., Japan).

2.13 - Electroelution of DNA

The construction of a partial genomic library required the isolation of a population of DNA fragments of a certain range of sizes. This was achieved by electroelution of restricted DNA. Genomic DNA (up to 100μg) was digested with a restriction enzyme then loaded onto a maxigel with markers. This required the casting of a 'slot gel' in which the plastic comb consists of a long continuous strip rather than separate teeth. After running, the markers were used to calculate the position on the gel of the required fragment, then a section of gel containing this fragment was cut out.

This was placed in large dialysis tubing (Scientific Industries International Inc., Loughborough, U.K.) containing 1 ml TE, then the tubing was sealed and placed back into the buffer tank. Electrophoresis was continued for 20 min, or until all DNA was seen, under UV light, to have left the gel slice. The gel slice was then removed and the DNA solution purified by two phenol / chloroform extractions and an ethanol / acetate precipitation. The pellet obtained was resuspended in TE.

2.14 - Dialysis

Dialysis of protein or DNA solutions against relevant buffers was an important step in various purification procedures. Dialysis tubing (Scientific Industries International) was prepared by boiling for 10 min in a solution containing 2% (w/v) sodium bicarbonate and 1mM EDTA pH 8. After rinsing in sterile distilled water, the tubing was boiled in 1mM EDTA solution pH 8, stored at 4°C until required, then rinsed again in sterile distilled water before use.

2.15 - Plasmids

The only type of plasmid used during this project was the high copy number phagemid pBluescript II KS+ (Stratagene). This plasmid contains an ampicillin resistance gene, an F1 origin of replication, the ColE1 origin of replication and a Lac Z gene incorporating a polylinker and a number of primer sites (Short *et al.*, 1988).

2.16 - Phosphatase treatment

Ligation of DNA fragments into plasmid vectors cut at a single restriction site can be increased in efficiency by preventing the plasmid self religating. This was done by treatment of the linearised vector with calf intestinal alkaline phosphatase (Amersham) according to the manufacturer's instructions. Subsequently the contaminating enzyme was removed by phenol / chloroform extraction and ethanol / acetate precipitation. The pellet of vector DNA was resuspended in TE and stored at -20^{o}C.

The phosphatase enzyme removes the phosphate groups from the 5' termini of a DNA molecule, making ligation, which requires 5' phosphate termini, impossible. However, if a DNA molecule still possessing 5' phosphate groups is introduced, it can be ligated to the phosphatase treated molecule.

2.17 - Ligations

Ligations of DNA with overhanging ends were carried out. Typical reactions required ;-

5ng of linearised vector, with insert in double molar excess
2µl of 10x ligation buffer (Gibco-BRL)
1µl of T4 DNA ligase (Gibco-BRL)
plus sterile distilled water to 20µl

Ligation was carried out for 12-18 h at 15^{o}C. Where required, for example during partial library construction, this reaction was scaled up.

2.18 - Production of competent cells

Competent cells were produced by rubidium chloride treatment . The strain of *E. coli* used was TG1, of phenotype supE hsdΔ5 thiΔ (lac-proAB) F' [traD36 proAB+ lacI lacZΔ M15] which is suitable for the selection of transformants by ampicillin selection and for blue / white selection of transformants containing insert. Competent cells were flash frozen and stored at -70°C.

2.19 - Transformation

Transformations were done by heat shock (Sambrook *et al.*, 1989). LB broth agar plates containing ampicillin at 200mg/ml and spread with 40µl of X-gal at 20mg/ml and 4µl of IPTG at 200mg/ml were used to select for transformants containing plasmid with insert DNA. Growth was allowed to proceed for 12-18 h at 37°C and 200 rpm in a Mk X incubator shaker (LH Engineering, Slough, Berkshire, U.K.).

2.20 - Screening genomic libraries

Various partial genomic libraries were constructed from *Sulfolobus* HT and LM. A specific set of DNA fragments was cloned into pBluescript vector, and transformants containing insert DNA found by blue / white selection. For each library approximately 1000 white colonies were picked and replated in duplicate on similar medium in large (23cm x 23cm) plates, together with positive controls where possible. These colonies were allowed to grow for 12-18 h at 37°C, then re picked onto positively charged nylon membrane (Boehringer) laid on top of similar fresh plates.

The colonies were allowed to grow for 12-18 h on the membrane, then the membrane was subjected to colony blotting procedure.

Colony blots were carried out according to the method of Grunstein and Hogness as published in the Boehringer DIG system manual. This results in the binding of DNA from the bacterial colonies to the filter. The filter was then prehybridised and hybridised as for Southern blots (see section 2.8).

2.21 - Radiolabelling oligonucleotides

Oligonucleotides were radiolabelled by end-labelling using polynucleotide kinase (Gibco-BRL). First the kinase buffer was made up, then used for the end labelling reaction with incubation at 37°C for 1 h. The labelled probe was subsequently used for hybridisation as detailed above, but with a different hybridisation buffer.

Kinase buffer -	Tris	0.5M
	$MgCl_2$	0.1M
	DTT	50mM
	Spermidine-HCl	1mM
	EDTA	1mM
	pH to 7.6	

End labelling reaction -	oligonucleotide	250ng
	kinase buffer	5µl (Gibco-BRL)
	T4 polynucleotide kinase	1µl + 1µl + 1µl (added every 20 min)
	γ^{32}P dATP (Amersham)	2.5µl (50µCi)
	to 50µl with sterile distilled water	

Hybridisation buffer -	200mM Sodium phosphate buffer pH 6.8	1ml
	20 x SSC	6ml
	Herring sperm DNA	250µl
	Denhardt's solution	1ml
	0.25M EDTA	80µl
	sterile distilled water	10.7ml

2.22 - Radiolabelling DNA by nick translation

DNA probes were also radioactively labelled by nick translation. This required ;

10µl DNA template (50ng)
2µl nucleotide mix (A,C,T only)
1µl DNA polymerase I (Gibco-BRL)
4µl NTB
1µl α-^{32}P dGTP (10µCi) (Amersham)
to 20µl with sterile distilled water

The above were incubated for 4 h at 15^{o}C, then checked for incorporation of radionuclide. This was done by loading the labelled probe onto a 20cm long column of 1g Sephadex G50 equilibrated in 10ml TE, eluting with TE and monitoring the fractions with a type 5.10 Geiger-Muller counter (Mini-Instruments, Burnham-on-Crouch, Essex, U.K.). The presence of two peaks of activity, the first corresponding to incorporated radionuclide, the second corresponding to unincorporated, confirmed that the labelling reaction had been successful. Care was taken during this procedure to minimise exposure to the radioactive reagents, necessitating the use of a perspex shield and

eppendorf carrier. The labelled probe was then denatured by heating to 95°C for 2min before adding to hybridisation mix. Hybridisation was carried out as previously described apart from the use of a different hybridisation solution;-

Hybridisation solution - 15ml 20x SSC
5ml 50x Denhardt's solution
2.5ml 10% (w/v) SDS
0.5 ml Herring sperm DNA, boiled before use
to 50ml with sterile distilled water.

2.23 - Plasmid DNA preparations

Four different methods of plasmid DNA purification were used. These were ;-

- Minipreps
- Maxipreps
- CsCl Minipreps
- Qiagen preparation

2.23a - Minipreps

Plasmids were prepared from a 2 ml for 12-18 h culture of colonies picked and grown in LB broth plus ampicillin (100µg/ml). The method used was alkaline lysis (Sambrook *et al.*, 1989). Plasmid DNA was stored at -20°C. This method gave sufficient DNA for tests such as restriction mapping, but not DNA of the purity required for sequencing.

2.23b - Maxipreps

Where large amounts of plasmid were required, 10 ml overnight cultures of colonies picked and grown in LB broth plus ampicillin (100μg/ml) were used to inoculate 500 ml cultures of similar media, allowed to grow for 12-18 h, then subjected to the large scale alkaline lysis method (Sambrook *et al.*, 1989) which includes a caesium chloride gradient ultracentrifugation step. This longer procedure yielded large quantities of DNA of a purity sufficient for sequencing and further genetic manipulation.

2.23c - Caesium chloride minipreps

Where high quality DNA was required, but only in relatively small amounts, from 2ml culture, caesium chloride minipreps were carried out (Saunders and Burke, 1990). This combines the usual alkaline lysis method with ethidium bromide precipitation of contaminating protein, and can be completed more quickly than a maxiprep.

2.23d - Qiagen preparations

Qiagen (Qiagen, Germany) preparation of plasmid DNA relies on the selective binding to then elution of DNA with immobilised glass beads under different saltconcentrations. Two kits were used according to the manufacturer's protocols. These were the Qiagen column kit and the Qiagen spin plasmid kit, which differ only in that the miniprep kit allows elution by centrifugation and is therefore faster and more convenient than the earlier column kit which relies on gravity elution.

2.24 - DNA purification from agarose gels

DNA fragments from agarose gels were purified either by electroelution, as detailed above (section 2.11), or by the use of the Qia-Ex kit (Qiagen) according to manufacturer's protocol. This is similar to the Qiagen procedure, but optimised for DNA contained in gel fragments.

2.25 - Manual plasmid sequencing

Sequencing of plasmid DNA was carried out by the dideoxy chain termination method (Sanger *et al.*, 1977) .

The polyacrylamide sequencing gel was constructed as follows ;-

Stock solutions

Urea solution - 233.5 g urea
100 ml 5x TBE (Sambrook *et al.*, 1989)
to 500 ml with sterile distilled water

Acrylamide solution - 96.5 g acrylamide
233.5 g urea
3.35 g bis-acrylamide
100 ml 5x TBE
to 500 ml with sterile distilled water

A Sequi-gel apparatus was used (Biorad, Hemel-Hempsted, Herts, U.K.)
One plate was coated with Repelcote before assembly of the apparatus.

Gel mix consisted of 68 ml acrylamide solution plus 102 ml urea solution. 30 ml of this mixture was taken to use as sealing gel at the base of the plates. This was polymerised using 150µl TEMED and 150µl of 25% (w/v) ammonium persulphate solution, and allowed to set for 5 min before pouring the main gel. The main gel consisted of the remainder of the gel mix polymerised with 140µl TEMED and 350µl of 25% (w/v) ammonium persulphate solution. This was poured, the comb inserted, and the gel allowed to set for 3 or more hours.

The gel was pre-run in 1x TBE for 1-2 h at 2.6 kV and 80 mA using a Consort E752 powerpack (Flowgen). When the temperature of the gel reached 45^{o}C, samples were loaded, then the gel was run for a further 2-8 h.

Sequencing reactions were carried out according to the manufacturer's protocol using Sequenase II kit (Amersham) and ^{35}S dGTP at 10µCi/µl (Amersham). 1-2µg of plasmid DNA were used for each reaction, with 50ng of the relevant primer.

After running, the gel was soaked in 10% (v/v) methanol, 10% (v/v) glacial acetic acid for 30 min, sandwiched between blotting paper (3M) and clingfilm then dried using a heated, vacuum, gel drier. When dry, the clingfilm was removed and the gel autoradiographed using X-ray film (Fuji RX), developer (Kodak) and fixer (Kodak).

2.26 - Automatic plasmid sequencing

Sample preparation for automatic sequencing consisted of mixing 1 μg of plasmid DNA with 3.2 pmoles primer in a final volume of 10.5 μl, made up with sterile distilled water. The automatic sequencing machine used was a model 373A (Applied Biosystems) and was run exclusively by Lesley Ward. Results were received in graphical format.

2.27- Protein Estimation

Estimation of protein concentration from whole cells was carried out using the method of Lowry et al (1951). Soluble protein concentrations were estimated using the Bradford method (1976)

.

2.28 - SDS-polyacrylamide gel protein electrophoresis

Solutions required

Tris / glycine electrode buffer	-	15.15 g tris
(pH 8.3)		72 g glycine
		5 g SDS
		plus distilled water to 5 l
Tris / SDS solution (pH 6.8)	-	30.27 g tris
		2 g SDS
		plus distilled water to 1 l

Tris / SDS solution (pH 8.8)	-	90.82 g tris 2 g SDS plus distilled water to 1 l
Acrylamide solution	-	30 % (w/v) acrylamide 0.8 %(w/v) N'-N-methylenebisacrylamide plus distilled water filtered then stored at 4°C
Lysis solution (pH 9)	-	50 mM Tris 1 mM EDTA
Sample buffer	-	25 ml Tris / SDS pH 6.8 2 g SDS 10 ml glycerol 5 ml 2-mercaptoethanol 0.1 ml 1% (w/v) bromophenol blue plus distilled water to 100 ml
Resolving gel (made freshly)	-	10 ml acrylamide solution 1.5 ml 1 % (w/v) ammonium persulphate 7.5 µl TEMED 15 ml Tris / SDS pH 8.8 3.5 ml distilled water
Stacking gel (made freshly)	-	1 ml acrylamide solution 5 ml Tris / SDS pH 6.8 4 ml 0.25% (w/v) ammonium persulphate 5 µl TEMED

Approximately 50 μg of biomass were washed in sterile distilled water, then resuspended in 50 μl lysis buffer. Alternatively, protein bands from native gels were excised and used as samples. Sample buffer was added and the resulting mixture boiled for 5 min, then microcentrifuged to remove insoluble debris. The SDS-PAGE (Sodium dodecylsulphate-polyacrylamide gel electrophoresis) with discontinuous buffers technique was carried out using an LKB 2001 vertical electrophoresis system (LKB, Bromma, Sweden). Gel plates were cleaned with acetone, then distilled water and air dried. Resolving gel was poured, overlaid with 1 ml distilled water and allowed to polymerise for 1 h. Stacking gel was then poured, and gel allowed to polymerise for 1 h. Electrophoresis was at 40-80 mA for 2-4 hours or 5 mA for 12-18 h using an LKB 2197 power pack. Low molecular weight protein standards (Pharmacia Inc., Uppsala, Sweden) were used giving bands at 94 kDa, 67 kDa, 43 kDa, 30 kDa, 20.1 kDa and 14.4 kDa. Gels were usually stained using either Coomassie R or silver nitrate staining.

2.29 - Semi-denaturing SDS polyacrylamide gel electrophoresis

Semi-denaturing polyacrylamide gels, (Francis and Becker, 1984), were run prior to certain haem stains. A method similar to that of SDS-PAGE was used but with the following differences ;-

Electrode buffer	0.05 M Tris
	0.06 M Boric acid
	0.001 M EDTA
	pH 8.3
Sample buffer	As for SDS-PAGE but omitting 2-mercaptoethanol, mixed at room temperature for 10 min with no boiling.

2.30 - Native gel protein electrophoresis

Solutions required

10 x electrode buffer (pH 8.3)- 6 g Tris
28.8 g glycine
plus distilled water to 1 l

Tris buffer (pH 8.9) - 36.6 g Tris
plus distilled water to 100 ml

Tris buffer (pH 6.7) - 5.98 g Tris
plus distilled water to 100 ml

Acrylamide solution - 30 % (w/v) acrylamide
0.8 %(w/v) N'-N-methylenebisacrylamide
plus distilled water
filtered then stored at 4^{o}C

Stacking acrylamide solution - 10 % (w/v) acrylamide
2.5 %(w/v) N'-N-methylenebisacrylamide
plus distilled water
filtered then stored at 4^{o}C

Native sample buffer - 25 ml Tris buffer pH 6.7
10 ml glycerol
0.1 ml 1% (w/v) bromophenol blue
plus distilled water to 100 ml

Resolving gel (made freshly) - 5 ml Tris buffer pH 8.9

10 ml acrylamide solution

1.5 ml 1% (w/v) ammonium persulphate

23.5 ml distilled water

Stacking gel (made freshly) - 2.5 ml Tris buffer pH 6.7

5 ml stacking acrylamide solution

1 ml 1% (w/v) ammonium persulphate

11.5 ml distilled water

Approximately 50 μg of bacteria were washed in sterile distilled water, then resuspended in 50 μl lysis buffer (see section 2.27). Native sample buffer was added. The native PAGE with discontinuous buffers technique was carried out using an LKB 2001 vertical electrophoresis system. Gel plates were cleaned with acetone, then distilled water and air dried. Resolving gel was poured, overlaid with 1 ml distilled water and allowed to polymerise for 1 h. Stacking gel was then poured, and gel allowed to polymerise for 1 h. Electrophoresis was at 40-80 mA for 2-4 hours or 5 mA for 12-18 h using an LKB 2001 power pack. High molecular weight protein standards (Pharmacia) were used giving bands at 669 kDa, 440 kDa, 232 kDa, 140 kDa, and 67 kDa. Gels were usually stained using either Coomassie R or silver nitrate staining.

2.31 - Protein gel staining

Protein gels were Coomassie stained in ;- 0.1% Coomassie brilliant blue R250
40 % (v/v) methanol
10 % (v/v) acetic acid
for 2-12 h

Gels were destained in ;- 40 % (v/v) methanol
10 % (v/v) glacial acetic acid
for 1-2 h with many changes of solution

Gels were silver stained by the method of Wray *et al* (1981).

2.32 - Blotting protein from gels to filters

Polyvinylidene difluoride (Immobilon-P) membrane (Millipore) was cut to the required size, washed in methanol, then washed in blotting buffer.

blotting buffer - 29.07 g Tris
14.64 g glycine
500 ml methanol
1.5 g SDS
plus distilled water to 5 l

Gel and blotting paper (3M) were washed in blotting buffer for 20 min, then arranged in an electroblotter (LKB) with the PVDF membrane between the gel and the anode, and the blotting paper surrounding them both. A current of 250 mA was applied from an LKB 2001 power pack for 2-3 hours. Buffer was circulated in a

cooling coil through ice by an electric pump. The membrane was then stained with Coomassie as above but destained in 60% (v/v) methanol / 10% (v/v) glacial acetic acid then washed in 60% (v/v) methanol then dried.

2.33 - N-terminal protein microsequencing

The protein band to be sequenced was blotted to PVDF membrane as above, sealed in plastic, then sent to the Department of Biochemistry at Leicester for N-terminal microsequencing.

2.34 - Probe design and oligonucleotide production

Degenerate oligonucleotide probes were designed using N-terminal protein data. Degeneracies were minimised as far as possible by the choice of amino acids and the known codon bias of *Sulfolobus*, a low G+C organism. Oligonucleotide probes were 20-mers with 32, 64, or 128 degeneracies. Oligonucleotides were synthesised in the Department by Lesley Ward using an Applied Biosystems DNA synthesiser.

2.35 - DNA and protein database searches

DNA and protein database searches were carried out using the GCG system (Program Manual for Wisconsin package version 8, Sept. 1994, Genetics Computer Group, Wisconsin, USA.) with Genbank, EMBL and Swissprot databases. Programs used were ;-

seqed - Sequence input and editing

map - DNA restriction mapping and translation to protein sequence

fasta - Database searching for homologous sequences

blast - Database searching for homologous sequences

codonfrequency- Codon usage calculation

motifs - Protein motif searches

pileup - Protein sequence alignment of homologous sequences

lineup - Manual sequence alignment correction

distances - Generation of distance matrices by uncorrected, Jukes Cantor and Kimura algorhythms

growtree - Plotting of phylogenetic trees using distance matrices by UPGMA and neighbour joining procedures

Additionally, dendrodrams were drawn using the Fitch program from a distance matrix derived using the Protdist program of PHYLIP (Felsenstein, 1995).

2.36 - Two dimensional protein gel electrophoresis

Approximately 50µg of bacteria were lysed by incubation at 37°C for 30 min with 50mM Tris / 10mM EDTA at pH 9. 10µg of sample were then subjected to isoelectric focusing, using the modified O'Farrell technique (Ames and Nikaido, 1976). Ampholines in the pH range 3.5 to 10 were used. IEF was carried out on an LKB 2001-600 tube gel apparatus. After equilibration in SDS-containing sample buffer, protein was run on a second dimension, consisting of the normal SDS polyacrylamide gel. The resulting gel was then stained using silver nitrate.

2.37 - *Sulfolobus* membrane preparations

Membrane preparation was performed at pH 7 and pH 2. Membrane fractions were prepared from samples of tetrathionate-grown and iron-grown *Sulfolobus* LM (see sections 2.3 and 2.4), then lysed using a French press (see section 2.5). The lysed cells were centrifuged at 7740 g using a Beckman JA-20 rotor. This centrifugation step was necessary to remove debris and unbroken cells. The lysate was subsequently ultracentrifuged at 120000 g in a Beckman SW-41 Ti rotor for 2 h at 4°C, producing a supernatant containing the soluble fraction of the cells, and a pellet consisting of the membrane fraction. The pellet was resuspended in 20 mM Tris-H_2SO_4 at pH 7 or pH 2 and ultracentrifuged again under the same conditions. This step was included to wash the membrane fraction.

In order to solublise membrane proteins an additional step was necessary. Three different membrane solublisation methods were attempted, using the detergents Nonidet P-40 (NP-40), sarkosyl, and MEGA 10.

NP-40

Membrane pellet was resuspended in sample buffer II (SB II), regularly mixed over the course of 1 h at 4°C, then ultracentrifuged for 1 h at 150000 g in a Beckman SW-41 Ti rotor.

SB II - 0.5% (v/v) NP-40
0.5% (w/v) betaine
10% (v/v) glycerol
1mM DTT
20 mM Tris (pH 7 or pH 2)

MEGA 10

Membrane pellet was washed with 20 mM Tris / 10mM EDTA pH 7.5, washed with 1% (w/v) sodium cholate, then resuspended in 1% (v/v) MEGA 10 / 2mM sodium phosphate buffer at pH 8 or pH 2, regularly mixed over the course of 1 h at 4°C, then ultracentrifuged for 1 h at 150000 g in a Beckman SW-41 Ti rotor (Wakagi et al., 1989).

Sarkosyl

Membrane pellet was resuspended in 50 mM malonate / 1mM EDTA pH 5.5, an equal volume of 2% (w/v) sarkosyl / 1M NaCl / 10mM Tris at pH 8 or pH 2 was added and regularly mixed over the course of 1 h at 4°C, then ultracentrifuged for 1 h at 150000 g in a Beckman SW-41 Ti rotor (Anemüller and Schäfer, 1989)

2.38 - Spectrophotometry

Oxidized-minus-reduced spectra of whole cells or cell fractions were produced using a Beckman DU 70 scanning spectrophotometer at wavelengths between 380 nm and 720 nm. Cells components were reduced using a few grains of sodium dithionite. Cell fractions were also scanned directly using only buffer as a blank.

2.39 - Column chromatography

A number of different chromatography columns were used in cytochrome purification attempts;-

2.39a - DEAE Sephacel

DEAE Sephacel is a weak anion exchange resin with a pH range of 2-9. An 8 cm x 0.5 cm conventional chromatography column was used. SB II (pH 7) was used to buffer the column, and dilutions of 1M NaCl / SB II (pH 7) were used for elution. Elution was carried out stepwise with 15 ml of 50, 100, 200, 300, 400, and 500 mM salt. The column was run at 4^{o}C. All buffers were filtered and de-gassed before use.

Additionally, DEAE Sephacel minicolumns were run in poly-prop disposable chromatography columns (Biorad). SB II (pH 7) was used to buffer the column, and dilutions of 1M KCl / SB II (pH 7) were used for elution. Elution was carried out stepwise with 1 ml of 100, 250, 500, and 1000 mM salt. Also elutions with 25, 50, 100 and 250 mM salt were tried.

2.39b - Q-Sepharose

Q-Sepharose is an ion exchange matrix. This column was run on an FPLC system (Pharmacia). Column volume was 7 ml. The buffer used was SB II (pH 7), elution was carried out using a 0-200 mM gradient of NaCl in SB II (pH 7). All buffers were filtered and de-gassed before use.

2.39c - Superose 12

Superose 12 was a size exclusion column run using the FPLC. The buffer used was SB II (pH 7). The column was washed using a 0-300 mM gradient of NaCl in SB II (pH 7). All buffers were filtered and de-gassed before use.

2.39d - Mono-Q HR 5/5

This was an anion binding column run on the FPLC. The buffer used was SB II (pH 7). Elution was carried out using a 0-1 M gradient of KCl in SB II (pH 7). All buffers were filtered and de-gassed before use.

2.39e - DEAE Cellulose

DEAE cellulose minicolumns were run in poly-prop disposable chromatography columns (Biorad). SB II (pH 7) was used to buffer the column, and dilutions of 1M KCl/ SB II (pH 7) were used for elution. Elution was carried out stepwise with 1 ml of 100, 250, 500, and 1000 mM salt.

2.39f - Sephacryl S200 and Sepharose CL6B

Two size exclusion matrices, Sephacryl S200 and Sepharose CL6B minicolumns were run in poly-prop disposable chromatography columns (Biorad). SB II (pH 7) was used to buffer and elute from the columns.

2.39g - Hydroxylapatite

Hydroxylapatite is a weak ion exchange resin with additional adsorption characteristics for some proteins at different salt concentrations and pH values. An 8 cm x 0.5 cm conventional chromatography column was used (Pharmacia). SB II (pH 7) was used to buffer the column, and dilutions of 1M potassium phosphate buffer / SB II (pH 7) were used for elution. Elution was carried out stepwise with 15 ml of 10, 50, 100, 150, and 200 mM phosphate. The column was run at 4°C. All buffers were filtered and de-gassed before use.

Additionally, hydroxylapatite minicolumns were run in poly-prop disposable chromatography columns (Biorad). 10 mM potassium phosphate (pH 7) / 0.05 % (w/v) sarkosyl was used to buffer the column, and dilutions of 1M potassium phosphate (pH 7) / 0.05 % sarkosyl were used for elution. Elution was carried out stepwise with 1 ml of 10, 200, 400, 500, 600, 700, 800 and 1000 mM phosphate.

The minicolumn method was scaled up to 5 ml chromatography columns (Pharmacia). Elution was carried out stepwise with 5 ml of 10, 200, 400, 500, 600 and 700 mM phosphate, and 20 ml of 1000 mM phosphate.

2.40 - Haem stains

Two different methods of staining polyacrylamide gels for proteins possessing haem groups were used.

2.40a - TMBZ haem stain

Native polyacrylamide gels were stained using 3,3',5,5'-tetramethylbenzidine (TMBZ) (Goodhew *et al.*, 1986) as follows;

Stain solution -	TMBZ	-	1.25 mM
	sodium acetate	-	200 mM
	methanol	-	30 % (v/v)

Stain solution was made up, added to the gel and agitated for 30 min at room temperature. Hydrogen peroxide was added to 26mM, then the gel was agitated for a further 15 min at room temperature. Subsequently it was washed twice in 200 mM sodium acetate / 30 % (v/v) propanol. Any photographs were carried out within 2 h.

2.40b - DMB haem stain

Native, semi-denaturing and SDS polyacrylamide gels were stained using 3,3'-dimethoxybenzidine (DMB) (Francis and Becker,1984). The gel was washed with 200 ml of 12.5 % (w/v) TCA for 30 min at room temperature, then with double distilled water for 30 min.

Stain solution -	DMB	200 mg
	hydrogen peroxide	400 µl of 30 % (v/v) solution
	sodium citrate buffer	20 ml of 0.5 M solution pH 4.4

The gel was stained using the above solution for 1 h at room temperature, then washed in water. Photography was carried out within 2 h.

2.41 - Sucrose gradient

A solution of 90 % (w/v) sucrose in 20 mM tris-H_2SO_4 pH 7 / 0.05 % (w/v) sarkosyl was made up. This was diluted with 20 mM tris-H_2SO_4 pH 7 / 0.05 % (w/v) to concentrations of 10, 15, 20, 25, 30, 35, 40, 50, 60, 70, 80 % sucrose. 1 ml layers of each concentration plus a layer of the 90 % stock, were carefully poured into a polypropylene tube, with the lowest concentration at the top. Sample was carefully overlaid and the gradient ultracentrifuged at 150000 g in Beckman SW-41 Ti for 18 h, being allowed to decelerate with no application of the brake. 1 ml samples were carefully removed.

2.42 - Isoelectric focusing

Isoelectric focusing of membrane proteins eluted from SDS-PAGE on slab polyacrylamide gels rather than tube gels was carried out according to the method of Rubin and Leonardi (1983). Ampholines of pI 5-7 and pI 3-10 and isoelectric focusing ladder markers were used. Gels were Coomassie stained.

RESULTS AND DISCUSSION

CHAPTER 3

The 16S rRNA gene of *Sulfolobus* strain HT

3.1 - Introduction

At present the primary technique of bacterial classification is the comparison of rRNA sequences (Ludwig and Schleifer, 1994). Of the genus *Sulfolobus*, only three representatives have been characterised by the sequencing of their 16S rRNA genes. These are *Sulfolobus solfataricus* (Olsen *et al.*, 1985), *Sulfolobus shibatae* (Grogan *et al.*, 1990) and *Sulfolobus acidocaldarius* (Kurosawa and Itoh., 1993). *S. acidocaldarius*, however, was found to possess an identical 16S rRNA sequence to that of *S. solfataricus*, probably due to contamination of early culture collection stocks (Zillig, 1993). More than ten different isolates of *Sulfolobus*-like archaea were available at Warwick, including species of *Acidianus*, *Desulfurolobus* and *Metallosphaera* as well as un-named isolates such as strains LM and HT. Clearly, to aid in the identification and classification of isolates 16S rRNA sequencing is desirable. As part of on-going research to sequence the 16S rRNA genes of these available isolates, the aim of this project was to isolate and sequence the 16S rRNA gene of *Sulfolobus*-like strain HT. This strain appears to possess the greatest potential of those compared at Warwick for mineral leaching, mainly due to its high temperature optimum of 85^{o}C.

3.2 - Previous work and project objectives

Strain HT had been grown at Warwick using pyrite and tetrathionate as substrates. A large portion of the gene coding for the 16S rRNA of strain HT had previously been cloned by J. Owen via PCR of genomic DNA using consensus primers 26f, specific for archaeal sequences, and 1492r, of ubiquitous rRNA homology (Lane, 1991). This clone was designated pHT 1, possessing the cloned 1.5 kb of strain HT 16S rRNA inserted with *Eco* RI termini into a TA cloning kit vector. In order to clone and sequence the full length 16S rRNA gene and gain some information on the 5'

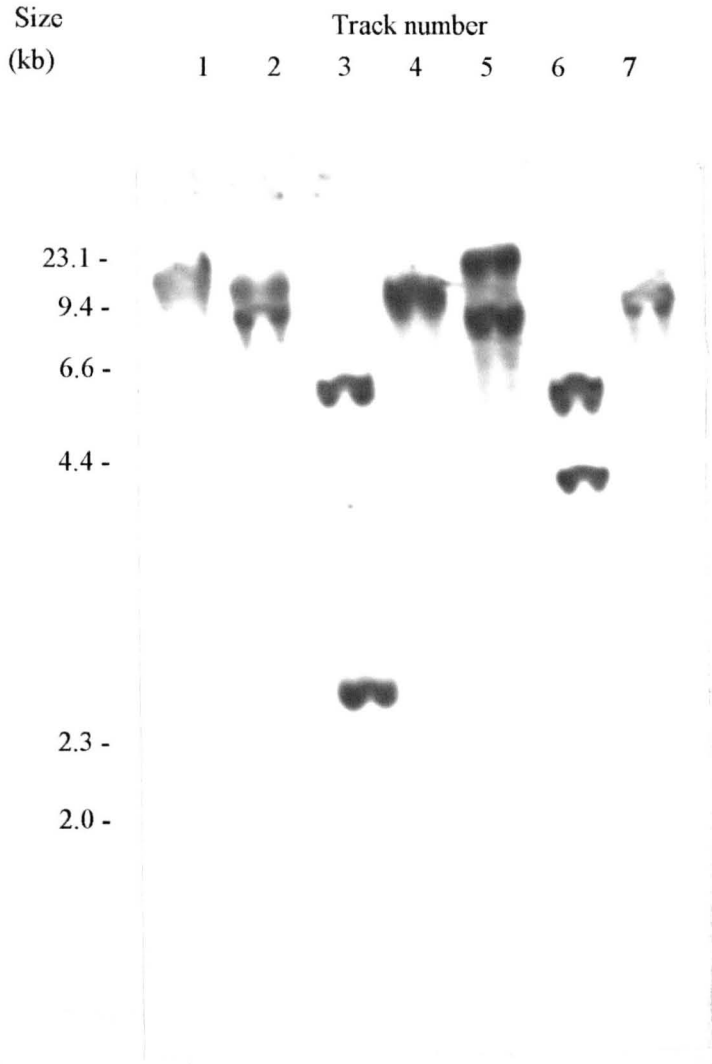
flanking region and intergenic region of the primary rRNA transcript, the isolation of a larger clone was necessary. This was to be isolated from a partial genomic library using the PCR product as a DNA probe.

3.3 - Detection of genes coding for the 16S rRNA in strain HT

Restriction digests of strain HT genomic DNA were carried out using a variety of endonucleases, *Sal* I, *Hind* III, *Eco* RV, *Eco* RI, *Pst* I and *Bam* HI. DNA samples from each of these digests, together with unrestricted genomic DNA, were separated by electrophoresis, Southern blotted and probed for the presence of the 16S rRNA gene. The cloned fragment from pHT 1 was used as a probe. This was excised by digestion with *Eco* RI, purified by the Qia-Ex procedure and non-radioactively labelled using the DIG DNA labelling kit. This probe was designated HT-1. Additionally the 26f oligonucleotide primer was used as a probe. This was non-radioactively labelled using the DIG oligonucleotide 3' end labelling kit. Hybridisations with the HT-1 probe were carried out at 65°C, with washes in 5 x SSC, 0.1 % SDS and 2 x SSC, 0.1 % SDS at 65°C. Hybridisations with the 26f oligonucleotide probe were carried out at 50°C, with washes in 5 x SSC, 0.1 % SDS and 2 x SSC, 0.1 % SDS at 50°C.

Luminescent detection showed probe HT-1 hybridisation to two discrete bands of restricted DNA in each restriction digest (see Fig. 3.1 overleaf). Hybridisation of oligonucleotide probe 26f gave an identical pattern. This demonstrated the presence of two copies of the 16S rRNA gene in the strain HT genome. The HT-1 probe was long enough to detect fragments of the gene with internal restrictions, but the 26f probe was too short (16 nucleotides) for this to have occurred. Multiple copies of rRNA genes are not uncommon, and the differences between them are usually insignificant due to frequent crossover events.

Figure 3.1 Autoluminescent photograph of probe HT-1 hybridisation at 65°C to restriction digests of strain HT genomic DNA. Tracks contain strain HT genomic DNA ; uncut (1), *Sal* I (2), *Hin*d III (3), *Eco* RV (4), *Eco* RI (5), *Pst* I (6) and *Bam* HI (7).



3.4 - Isolation of a clone homologous to probe HT-1

From the previous hybridisation experiment, it was seen that two *Hin*d III restriction fragments, one of approximately 6.8 kb and one of approximately 2.7 kb showed hybridisation to probe HT-1 (see Fig. 3.1). Therefore two partial libraries of strain HT genomic DNA fragments of 1 kb regions around 6.8 kb and around 2.7 kb were constructed. The fragments were formed by restriction with *Hin*d III then eluted from an agarose gel. A pBluescript KS+ phagemid vector was chosen as the cloning vector. This is a high copy number vector which has a large polylinker and ampicillin resistance and is amenable to blue / white selection of transformants with inserts (Short *et al.*, 1988). The vector was restricted with *Hin*d III, phosphatased to prevent self religation and repurified. After purification, the genomic DNA fragments were ligated into the vector, then transformed into *E. coli* strain TG 1. White colonies were selected and replated in duplicate. *E.coli* strain TG 1 containing pHT 1 was used as a positive control. One set of colonies was colony blotted and hybridised with probe HT-1 at 60°C with washes in 5x and 2x SSC at the same temperature. Initially, difficulties with the non-radioactive probing of colony blots were encountered. A *Pst* I partial library was constructed and probed, with a similar outcome. However, by remaking the *Hin*d III partial libraries and radioactively labelling probe HT-1 by nick translation, 1800 colonies of a 5-8 kb *Hin*d III genomic library were screened, providing two positive clones, designated 16SA and 16SB (see Fig. 3.2 overleaf).

Both positive clones were cultured and plasmid DNA prepared by miniprep procedure. These were restricted using *Hin*d III and using *Pst* I. The *Hin*d III digests gave bands of 6.5 kb and 3.0 kb in both cases, as expected. The *Pst* I digests gave bands of 4 kb, 2 kb, 1.8 kb, 1.2 kb and 0.5 kb for both positive clones indicating that both were identical, and had been inserted into the vector in the same orientation.

Figure 3.2 Autoradiograph of a colony blot of a 5-8 kb *Hind* III partial library of strain HT genomic DNA hybridised with probe HT-1

3.5 - Sequencing of clone 16SA

Plasmid DNA was prepared from clone 16SA by maxiprep and Qiagen procedures. DNA plasmid sequencing was carried out using both manual and automatic sequencing techniques. Ideally, the entire sequence of both DNA strands would have been elucidated. However, time was only available to sequence the majority of both strands which could be assembled into the primary DNA sequence of the whole fragment (see Fig. 3.3 overleaf). This may have lead to minor sequence inaccuracies. Primers used were consensus 16S rRNA primers (Lane, 1991).

These were (using the usual *E.coli* numbering system) ;-

26f 127f 355f 499f 915f 1114f 1203f 1492f

110r 482r 896r 1099r 1492r

As well as the entire length of the 16S rRNA gene, elements both upstream and downstream were sequenced. 53 bases were sequenced upstream before a *Hin*d III site was found with which the fragment had obviously been inserted into the vector, as pBluescript sequence was found beyond this site. This was disappointing as a greater length of sequence could have provided more information on helices 5' to the 16S precursor processing stem. Downstream an additional 180 bases were sequenced. This gave the sequence of the intragenic region up to the beginning of the 23S precursor processing stem.

Figure 3.3 Diagram of clone 16SA, with *Hind* III site of insertion into the vector pBluescript (H) showing the extent of the 16S rRNA gene (1498 bases). Arrows represent the extent and direction of sequences obtained. Drawn to scale (1781 bases total length)

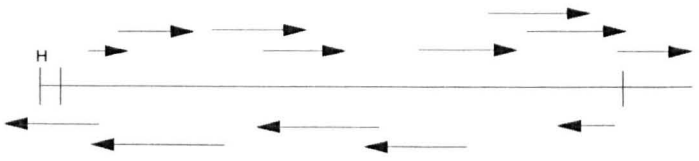


Figure 3.4 DNA sequence of the 16S rRNA gene and surrounding sequences of strain HT. Sequence is shown as for DNA with deoxythymidine (T), whereas in rRNA these positions would be replaced by uridine (U).

5' flanking region

```
                                                        -53  AAG
-50  CTTACTTCTG ACAGGGGAAA CCGAGAGTGG TGTAAGCTCT CCTGCCCCCA
```

16S rRNA gene sequence

```
  1  ATTCCGGTTG ATCCTGCCGG ACCCGACCGC TATGGGGGTA GGGCTAAGCC
 51  ATGGGAGTCG TACGCCCTCG GGTAAGAGGG CGTGGGGGAC GGCTGAGTAA
101  CACGTGGCTA ACCTACCCTC GGGACCCGGA TAACTCCGGG AAACTGGAGC
151  TAATCCGGGG CAGGCGAAGG GTACTGGAAC GTCCCTTCGC CTAAAGGGGT
201  ATGGGCTATT TCCCGCTCAT GCCCGCCCGA GGATGGGGCT GCGGCCTATC
251  AGGCTGTTGG CGGGGTAACG GCCCGCCAAA CCGATAACGG GTAGGGGCCG
```

301 TGAGAGCGGG AGCCCCCAGT TGGGCACTGA GACAAGGGCC CAGGCCCTAC

351 GGGGCGCACC AGGCGCGAAA CGTCCCCAAT GCGGGAAACC GTGAGGGCGC

401 TACCCCCAGT GCTCCCGAAA GGGAGCTTTT CCCCGCTTTA GAACGGCGGG

451 GAATAAGCGG GGGGCAAGAC TGGTGTCAGC CGCCGCGGTA ATACCAGCCC

501 CGCGAGTGGT CGGGACGCTT ACTGGGCTTA AAGCGCCCGT AGCCGGCCCT

551 GCAAGTCACT GCTTAAAGAC CCCGGCTCAA CCCGGGAAAG GGCAGTGATA

601 CTGCAGGGCT AGGGGGCGGG AGAGGTCGGA GGTACTCCCG GAGTAGGGGC

651 GAAATCCACA GATCCCGGGA GGACCACCAG TGGCGAAAGC GTCCGGCCAG

701 AACGCGCCCG ACGGTGAGGG GCGAAAGCCG GGGTAGCAAA AGGGATTAGA

751 TACCCCTGTA GTCCCGGCTG TAAACGATGC AGGCTAGGTG TCGCGTAGGC

801 TTTGCGCCTA CGCGGTGCCG CAGGGAAACC GGTAAGCCTG CCGCCTGGGG

851 AGTACGCCCC GAAGGGTGAA ACTTAAAGGA ATTGGCGGGG GAGCACCACA

901 AGGGGTGGAA CCTGCGGCTC AATTGGAGTC AACGCCTGGA ATCTCACCGG

951 GGGAGACCGC AGGATGACGG CCAGGCTAAC GACCTTGCCA GACTCGCGGA

1001 GAGGAGGTGC ATGGCCGTCG CCAGCTCGTG TTGTGAAATG TCCGGTTAAG

1051 TCCGGCAACG AGCGAGACCC CCACCCTTAG TTGGTATCCC GGTCTCCGGA

1101 CCGGGACCAC ACTAAGGGGA CTGCCGGCGT AAGCCGGAGG AAGGAGGGGG

1151 CCACGGCAGG TCAGCATGCC CCGAAACCCC CGGGCCGCAC GCGGGTTACA

1201 ATGGCAGGGA CAGCGGGATT CCGACCCCGA AAGGGGGAGG CAATCCCTCA

1251 AACCCTGCCT CAGTTGGGAT CGAGGGCTGA AACTCGCCCT CGTGAACGAG

1301 GAATCCCTAG TAACCGCACG TCAACAACGT GCGGTGAATA CGTCCCTGCT

1351 CCTTGCACAC ACCGCCCGTC GCTCCACCCG AGTGAGGAAG AAGTGAGGCT

1401 CCTTGCCCTT CGGGGTGGGG AGTCGAACTT CTTCCTCGCG AGGGGGGAGA

1451 AAGTCGTAAC AAGGTAGCCG TAGGGGAACC TGCGGCTGGA TCACCTCA

Intragenic region

1499 CATTCAACAC TGTTGGTGCT CCCCTAGAGG GGCAGGAGAG CCACTTAAAC

1549 TCTCGGCTTC CCCATTTACT TAATGCGGCT CCACTCTGTG GAGTGGAGCT

1599 AAGAGCCTAG GACTCGAGTT TAGCAGCTCG AGTCCAGTGA GGCTAGCCAA

1649 CGACCAGAGT TAAACAGTCA CCTAGGGACC

Figure 3.5 Alignment of the 16S rRNA gene sequence of strain HT (HT) with those of *Sulfolobus shibatae* (SSHIB) and *Sulfolobus acidocaldarius \ solfataricus* (SSOLF). Numbering is as for the strain HT gene. Differences are shown by asterisks.

```
        1
SSOLF   ATTCCGGTTG ATCCTGCCGG ACCCGACCGC TATCGGGGTA GGGATAAGCC
   HT   ATTCCGGTTG ATCCTGCCGG ACCCGACCGC TATGGGGGTA GGGCTAAGCC
SSHIB   AATCCGGTTG ATCCTGCCGG ACCCGACCGC TATCGGGGTG GGGCTAAGCC
                                                 *          *

        51
SSOLF   ATGGGAGTCT TACACTCCCG GGTAAGGGAG TGTGGCGGAC GGCTGAGTAA
   HT   ATGGGAGTCG TACGCCCTCG GGTAAGAGGG CGTGGGGGAC GGCTGAGTAA
SSHIB   ATGGGAGTCG TACGCTCCCG GGCAAGGGAG CGTGGCGGAC GGCTGAGTAA
                  * * *     *   * *  *    *

        101
SSOLF   CACGTGGCTA ACCTACCCTC GGGACGGGGA TAACCCCGGG AAACTGGGGA
   HT   CACGTGGCTA ACCTACCCTC GGGACCCGGA TAACTCCGGG AAACTGGAGC
SSHIB   CACGTGGCTA ACCTACCCTG AGGAGGGAGA TAACCCCGGG AAACTGGGGA
                        * *   ****       *             * *

         151
SSOLF   TAATCCCCGA TAGGGAAGGA GTCCTGGAAT GGTTCCTTCC CTAAAGGGCT
   HT   TAATCCGGGG CAGGCGAAGG GTACTGGAAC GTCCCTTCGC CTAAAGGGGT
SSHIB   TAATCTCCCA TAGGCGAGGA GTCCTGGAAC GGTTCCTCGC TGAAAGGTTC
           ***** *   ** * *   *      *  *** * **  **     ***

        201
SSOLF   ATAGGCTATT TCCCGTTTGT AGCCGCCCGA GGATGGGGCT ACGGCCCATC
   HT   ATGGGCTATT TCCCGCTCAT GCCCGCCCGA GGATGGGGCT GCGGCCTATC
SSHIB   ATGGGCTATT TCCCGCTCAT GAGCGCCTCA GGATGGGGCT GCGGCCCATC
          *              * **  ***    **               *    *

        251
SSOLF   AGGCTGTCGG TGGGGTAAAG GCCCACCGAA CCTATAACGG GTAGGGGCCG
   HT   AGGCTGTTGG CGGGGTAACG GCCCGCCAAA CCGATAACGG GTAGGGGCCG
SSHIB   AGGTAGTTGG GGGGGTAAGG GCCCCCCAAG CCTATAACGG GTAGGGGCCG
           **  *   *       *      *  * *   *

        301
SSOLF   TGGAAGCGGG AGCCTCCAGT TGGGCACTGA GACAAGGGCC CAGGCCCTAC
   HT   TGAGAGCGGG AGCCCCCAGT TGGGCACTGA GACAAGGGCC CAGGCCCTAC
SSHIB   TGAGAGCGGG AGCCCCCAGT TGGGCACTGA GACAAGGGCC CAGG.CCTAC
          **           *                                 *

        351
SSOLF   GGGGCGCACC AGGCGCGAAA CGTCCCCAAT GCGCGAAAGC GTGAGGGCGC
   HT   GGGGCGCACC AGGCGCGAAA CGTCCCCAAT GCGGGAAACC GTGAGGGCGC
SSHIB   GGGGCGCACC AGGCGCGAAA CGTCCCCAAT GCGCGGAAGC GTGAGGGCGC
                                          * *  *

        401
SSOLF   TACCCCGAGT GCCTCCGCAA GGAGGCTTTT CCCCGCTCTA AAAAGGCGGG
   HT   TACCCCCAGT GCTCCCGAAA GGGAGCTTTT CCCCGCTTTA GAACGGC.GG
SSHIB   CACCCCGAGT GCTCCCGTAA GGGAGCTTTT CCCCGCTCTA CAAAGGCGGG
        *     *      **   *     **                *   *  *   *
```

```
        450
SSOLF   GGAATAAGCG GGGGGCAAGT CTGGTGTCAG CCGCCGCGGT AATACCAGCT
   HT   GGAATAAGCG GGGGGCAAGA CTGGTGTCAG CCGCCGCGGT AATACCAGCC
SSHIB   GGAATAAGCG GGGGGCAAGT CTGGTGTCAG CCGCCGCGGT AATACCAGCC
                              *                                *

        500
SSOLF   CCGCGAGTGG TCGGGGTGAT TACTGGGCCT AAAGCGCCTG TAGCCGGCCC
   HT   CCGCGAGTGG TCGGGACGCT TACTGGGCTT AAAGCGCCCG TAGCCGGCCC
SSHIB   CCGCGAGTGG TCGGGACTCT TACTGGGCCT AAAGCGCCCG TAGCCGGCCC
                         ****                     *

        550
SSOLF   ACCAAGTCGC CCCTTAAAGT CCCCGGCTCA ACCGGGGAAC TGGGGGCGAT
   HT   TGCAAGTCAC TGCTTAAAGA CCCCGGCTCA ACCCGGGAAA GGGCAGTGAT
SSHIB   GACAAGTCAC TCCTTAAAGA CCTCGGCTCA ACCGGGGGAA TGGGGGTGAT
        **      *  **       *   *           *   * * *  ****

        600
SSOLF   ACTGGTGGGC TAGGGGGCGG GAGAGGCGGG GGGTACTCCC GGAGTAGGGG
   HT   ACTGCAGGGC TAGGGGGCGG GAGAGGTCGG AGGTACTCCC GGAGTAGGGG
SSHIB   ACTGTCGGGC TAGGGGGCGG GAGAGGCCAG CGGTACTCCC GGAGTAGGGG
           **                        ***  *

        650
SSOLF   CGAAATCCTT AGATACCGGG AGGACCACCA GTGGCGGAAG CGCCCCGCTA
   HT   CGAAATCCAC AGATCCCGGG AGGACCACCA GTGGCGAAAG CGTCCGGCCA
SSHIB   CGAAATCCTC AGATCTCGGG AGGACCACCA GTGGCGAAAG CGGCTGGCTA
              **     **                       *     * **  *

        700
SSOLF   GAACGCGCCC GACGGTGAGA GGCGAAAGCC GGGGCAGCAA ACGGGATTAG
   HT   GAACGCGCCC GACGGTGAGG GGCGAAAGCC GGGGTAGCAA AAGGGATTAG
SSHIB   GAACGCGCCC GACGGTGAGG GGCGAAAGCC GGGGCAGCAA AAGGGATTAG
                              *                 *       *

        750
SSOLF   ATACCCCGGT AGTCCCGGCT GTAAACGATG CGGGCTAGGT GTCGAGTAGG
   HT   ATACCCCTGT AGTCCCGGCT GTAAACGATG CAGGCTAGGT GTCGCGTAGG
SSHIB   ATACCCCTGT AGTCCCGGCT GTAAACGATG CAGGCTAGGT GTCACATGGG
              *                           *             *** *

        800
SSOLF   CTTAGAGCCT ACTCGGTGCC GCAGGGAAGC CGTTAAGCCC GCCGCCTGGG
   HT   CTTTGCGCCT ACGCGGTGCC GCAGGGAAAC CGGTAAGCCT GCCGCCTGGG
SSHIB   CTTAGAGCCC ATGTGGTGCC GCAGGGAAGC CGTTAAGCCT GCCGCCTGGG
           * *   *  ***                  *    *

        850
SSOLF   GAGTACGGTC GCAAGACTGA AACTTAAAGG AATTGGCGGG GGAGCACCAC
   HT   GAGTACGCCC CGAAGGGTGA AACTTAAAGG AATTGGCGGG GGAGCACCAC
SSHIB   GAGTACGGTC GCAAGACTGA AACTTAAAGG AATTGGCGGG GGAGCACCAC
              *   **   **

        900
SSOLF   AAGGGGTGGA ACCTGCGGCT CAATTGGAGT CAACGCCTGG AATCTTACCG
   HT   AAGGGGTGGA ACCTGCGGCT CAATTGGAGT CAACGCCTGG AATCTCACCG
SSHIB   AAGGGGTGGA ACCTGCGGCT CAATTGGAGT CAACGCCTGG AATCTTACTA
                                                     *  **
```

```
        950
SSOLF   GGGGAGACCG CAGTATGACG GCCAGGCTAA CGACCTTGCC TGACTCGCGG
   HT   GGGGAGACCG CAGGATGACG GCCAGGCTAA CGACCTTGCC AGACTCGCGG
SSHIB   GGGGAGACCG CAGGATGACG GCCAGGCTAA CGACCTTGCC TGACTCGCGG
                     *                                  *

        1000
SSOLF   AGAGGAGGTG CATGGCCGTC GCCAGCTCGT GTTGTGAAAT GTCCGGTTAA
   HT   AGAGGAGGTG CATGGCCGTC GCCAGCTCGT GTTGTGAAAT GTCCGGTTAA
SSHIB   AGAGGAGGTG CATGGCCGTC GCCAGCTCGT GTTGTGAAAT GTCCTGTTAA
                                                       *

        1050
SSOLF   GTCCGGCAAC GAGCGAGACC CCCACCCCTA GTTGGTATTC TGGACTCCGG
   HT   GTCCGGCAAC GAGCGAGACC CCCACCCTTA GTTGGTATCC CGGTCTCCGG
SSHIB   GTCAGGCAAC GAGCGAGACC CCCACCATTA GTTGGTATCC TGGTCTCCGG
           *                           **           *  *  *

        1100
SSOLF   TCCAGAACCA CACTAGGGGG ACTGCCGGCG TAAGCCGGAG GAAGGAGGGG
   HT   ACCGGGACCA CACTAAGGGG ACTGCCGGCG TAAGCCGGAG GAAGGAGGGG
SSHIB   GCCGGGACCA CACTAGTGGG ACTGCCGGCG TAAGCCGGAG GAAGGAGGGG
        *  * *           **

        1150
SSOLF   GCCACGGCAG GTCAGCATGC CCCGAAACTC CCGGGCCGCA CGCGGGTTAC
   HT   GCCACGGCAG GTCAGCATGC CCCGAAACCC CCGGGCCGCA CGCGGGTTAC
SSHIB   GCCACGGCAG GTCAGCATGC CCCGAAACCC CTGGGCCGCA CGCGGGTTAC
                                       *   *

        1200
SSOLF   AATGGCAGGG ACAACGGGA. TGCTACCTCG AAAGGGGGAG CCAAT.CCTT
   HT   AATGGCAGGG ACAGCGGGAT TCCGACCCCG AAAGGGGGAG GCAATCCCTC
SSHIB   AATGGCAGGG ACAGCGGGAT TCCGACCCCG AAAGGGGAAG GTAATCCCTT
                     *     *  * *   *          *   **   *   *

        1250
SSOLF   AAACCCTGCC GCAGTTGGGA TCGAGGGCTG AAACCCGCCC TCGTGAACGA
   HT   AAACCCTGCC TCAGTTGGGA TCGAGGGCTG AAACTCGCCC TCGTGAACGA
SSHIB   AAACCCTGCC GCAGTTGGGA TCGAGGGCTG AAACTCGCCC TCGTGAACGA
                  *                           *

        1300
SSOLF   GGAATCCCTA GTAACCGCGG GTCAACAACC CGCGGTGAAT ACGTCCCTGC
   HT   GGAATCCCTA GTAACCGCAC GTCAACAACG TGCGGTGAAT ACGTCCCTGC
SSHIB   GGAATCCCTA GTAACCGCAG ATCAACAATC TGCGGTGAAT ACGTCCCTGC
                         ** *          *

        1350
SSOLF   TCCTTGCACA CACCGCCCGT CGCTCCACCC GAGCGCGAAA GGGGTGAGGT
   HT   TCCTTGCACA CACCGCCCGT CGCTCCACCC GAGTGAGGAA GAAGTGAGGC
SSHIB   TCCTTGCACA CACCGCCCGT CGCTCCACCC GAGTAGGAGA GGGGTGAGGC
                                          *** **   **      *

        1400
SSOLF   CCCTTGCGAT ...AAGTGGG GGATCGAACT CCTTTCCCGC GAGGGGGGAG
   HT   TCCTTGCCCT TCGGGGTGGG GAGTCGAACT CCTTTCCCGC GAGGGGGGAG
SSHIB   CCCTTGCCTT T..AGGTGGG GGGTCGAGCT CCTTCCTCGC GAGGGGGGAG
        *      **  *****       **    *       * *

        1450
SSOLF   .AAGTCGTAA CAAGGTAGCC GTAGGGGAAC CTGCGGCTGG ATCACCTCA
   HT   AAAGTCGTAA CAAGGTAGCC GTAGGGGAAC CTGCGGCTGG ATCACCTCA
SSHIB   .AAGTCGTAA CAAGGTAGCC GTAGGGGAAC CTGCGGCTGG ATCACCTCA
```

3.6 - Phylogenetic analysis of the 16S rRNA gene of strain HT

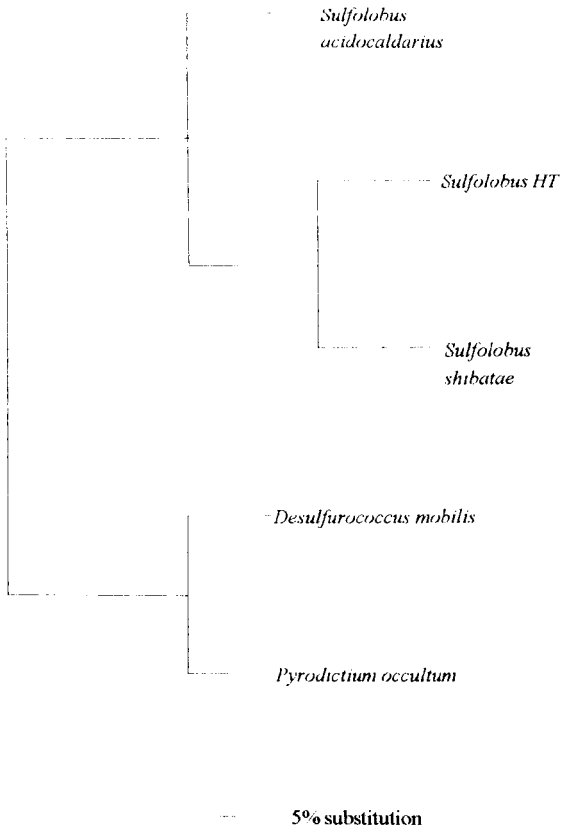
Distance matrices were calculated for sequences of 16S rRNA genes of *Sulfolobus solfataricus \ acidocaldarius*, *Sulfolobus shibatae*, strain HT and a range of other crenarchaeota using the Jukes-Cantor method (see Fig. 3.6). These showed strain HT segregating within the *Sulfolobus* taxon, justifying its designation as *Sulfolobus* HT. The closest group to that of the *Sulfolobus* organisms was that containing *Pyrodictium occultum* (Accession no. M21087) and *Desulfurococcus mobilis* (Accession no. M36474), as shown previously (Olsen *et al.*, 1994). It is interesting that *P. occultum* and *D. mobilis* are assigned to different genera whereas *S. acidocaldarius* and *S. shibatae* are assigned to the same genus despite their 16S rRNA sequences differing to a greater extent. With the continued sequencing of 16S rRNA from a wider range of archaea, the nomenclature of these organisms may require some reconsideration.

Within the genus *Sulfolobus*, the HT organism was found to be more closely related to *S. shibatae* than to *S. solfataricus \ acidocaldarius*. Using this matrix, a phylogenetic tree was produced using the UPGMA method (see Fig 3.7 overleaf).

Figure 3.6 Matrix of difference scores between some archaeal 16S rRNA genes expressed as substitutions per 100 bases.

	S.solf / acid	strain HT	*S. shibatae*	*D. mobilis*	*P. occultum*
S. solf / acid	0.00	10.76	9.47	13.72	14.03
strain HT		0.00	9.20	14.36	14.30
S. shibatae			0.00	14.36	14.30
D. mobilis				0.00	7.69
P. occultum					0.00

Figure 3.7 Phylogenetic tree of 16S rRNA genes from *Sulfolobus*, *Desulfurococcus* and *Pyrodictium* species, produced from a Jukes-Cantor distance matrix using UPGMA method.



From previous work including SDS-PAGE protein profiles, chromosomal DNA restriction enzyme patterns and mol % GC content values (Norris et al., pers.comm.) it appears that *Sulfolobus* isolates LM and BC are identical and probably of the same species as *Sulfolobus metallicus*. These isolates are thought to be more similar to *Sulfolobus* HT than to either *S. shibatae* or *S. acidocaldarius*.

The sulphur-oxidising strain of *S. solfataricus*, for which as yet no 16S rRNA data has been published appears to group with *S. shibatae*. *S. acidocaldarius* forms the third group. Another group may be represented by *Metallosphaera sedula* and novel *Sulfurococcus* isolates which both possess similar mol % GC content values.

3.7 - 16S rRNA structure of *Sulfolobus* HT

The primary sequence of *Sulfolobus* HT 16S rRNA (see Figs. 3.4 and 3.5) was compared with the secondary 16S rRNA structural model of *S. acidocaldarius*, published for *S. solfataricus* (Olsen *et al.*, 1985). The sequence conforms with that of *S. acidocaldarius* by possessing a higher proportion of canonical base pairs and fewer helical irregularities than 16S rRNA from mesophilic archaea, as expected from a thermophilic organism. Most differences between the two thermophilic sequences were neutral in terms of sequence complementarity and therefore in terms of predicted thermostability. There were a similar number of substitutions producing greater complementarity as producing lower complementarity, also predicting little difference in rRNA thermostability.

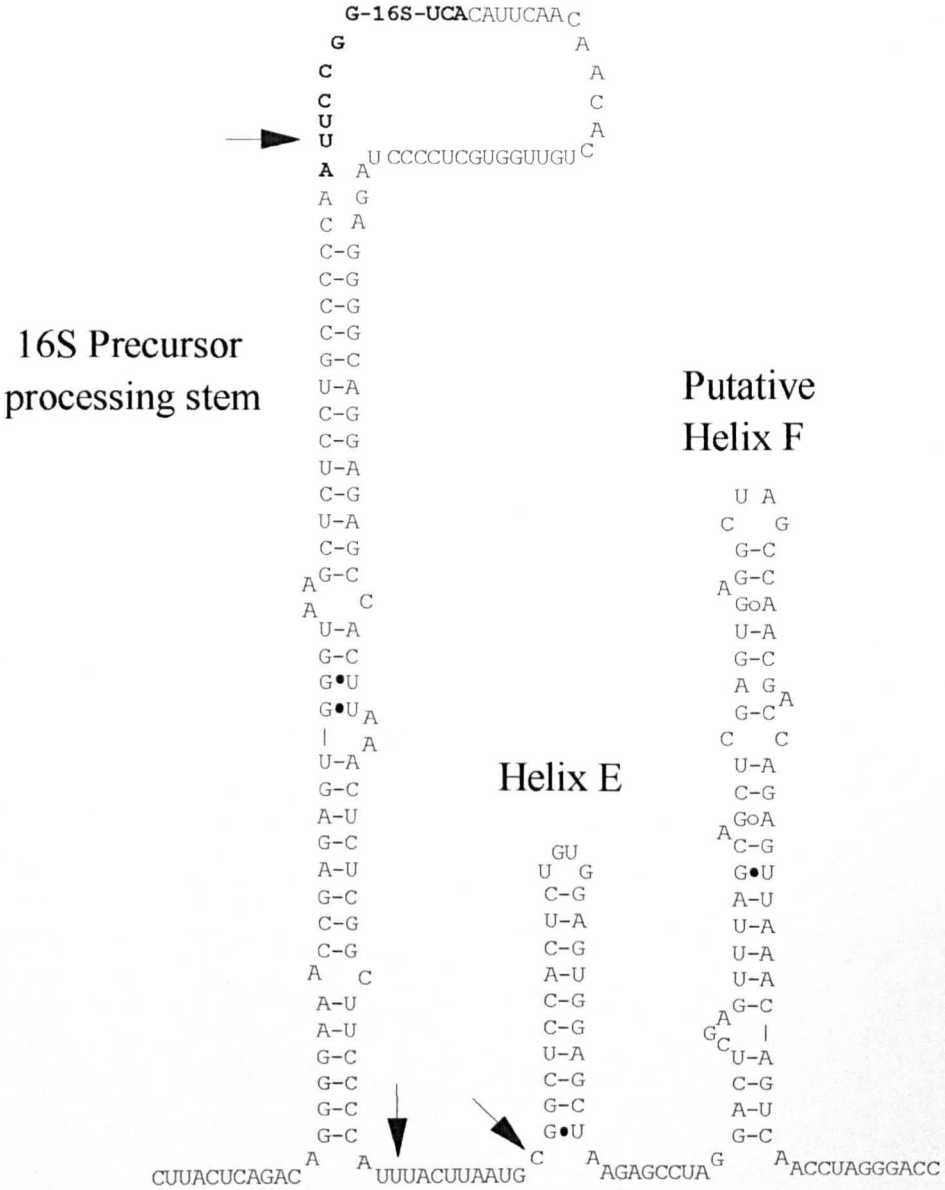
3.8 - Secondary structure of 5' and intergenic regions

The crenarchaeota, including *S. acidocaldarius*, and *S. shibatae* possess no tRNA genes in their intergenic spacer regions between 16S rRNA and 23S rRNA genes. This is in contrast to the *Halobacteria* and *Methanobacteria* (Brown *et al.*, 1989). In *S. acidocaldarius* the mature 16S and 23S rRNA are generated by cleavage of a 5 kb transcript. The 23S rRNA is cleaved from the 5 kb transcript by the typical bulge-helix-bulge motif found in most archaea, but the 16S rRNA is not, cleavage being found to occur at a number of novel restriction sites (Durovic and Dennis, 1994).

The *Sulfolobus* HT 5' flanking and intergenic regions sequenced were of low homology to those of *S. acidocaldarius* in comparison with the region coding for the mature 16S rRNA. However, secondary structure predictions provided evidence of a similar 16S precursor processing stem, intragenic 'helix E' structure and possibly the intragenic 'helix F' structure (see Fig. 3.8 overleaf). In addition, three of four of the sites identified as being restricted by novel endonucleases in *S. acidocaldarius* were conserved. These comparisons suggested that *Sulfolobus* HT follows a pathway for excision and processing of 16S rRNA from the primary rRNA operon transcript similar but not identical to that of *S. acidocaldarius*.

In summary, therefore, the 16S rRNA gene of strain HT shows structural characteristics expected from archaeal thermoacidophiles, both in its primary sequence and in the predicted secondary structure of flanking regions. Initial phylogenetic analysis of this sequence has suggested that strain HT should be included within the genus *Sulfolobus*, showing greater similarity to the chemolithotrophic *S. shibatae* than to the heterotrophic *S. acidocaldarius*.

Figure 3.8 Possible secondary structure for the 16S 5' flanking and intergenic regions of the primary rRNA transcript of Sulfolobus HT. The 5' and 3' ends of the mature 16S rRNA are shown in bold type. Arrows show cleavage sites for possible endonucleases conserved from *S. acidocaldarius*.



CHAPTER 4

Investigation of proteins induced by CO_2 limitation in *Sulfolobus* strain LM

4.1 - Introduction

Previous work (Norris *et al*, 1989), has demonstrated a correlation between CO_2 fixation by *Sulfolobus* strain LM and a protein of 330 kDa molecular weight, such fixation being dependant upon the presence of ATP and acetyl-CoA. The further investigation of this protein was continued using the same strain of *Sulfolobus*.

These studies aimed for a further characterisation of this protein via the cloning of the gene encoding its amino acid sequence. This necessitated the repetition of previous work to identify the protein and subsequently its partial purification. These steps were to be followed by the N-terminal amino acid sequencing of the protein and an attempt at cloning and sequencing the gene responsible for its production.

4.2 - Growth of *Sulfolobus* LM with differing CO_2 concentrations

Comparing cells grown under CO_2 limitation, for example those gassed with air, with those grown using excess CO_2 was the strategy used to identify proteins which may be involved in CO_2 fixation.

Separate cultures of *Sulfolobus* LM were grown on tetrathionate with 5% (v/v) CO_2 in air, 1% (v/v) CO_2 in air and in air. The rate of pyrite oxidation by *Sulfolobus* LM was reduced by 26% using air when compared with growth using 5% (v/v) CO_2 in air (Norris *et al*, 1989).

4.3 - SDS-PAGE of proteins from cells grown with CO_2 limitation and CO_2 excess

SDS-PAGE was carried out using samples of *Sulfolobus* LM grown under the above conditions. A high percentage (15% w/v) acrylamide gel was used to visualise protein bands of under 30 kDa molecular weight, whereas a medium percentage (10% w/v) acrylamide gel was used to visualise protein bands of over 30 kDa molecular weight. Gels were examined for the presence of increased proportions of protein bands in the samples grown under air ; i.e. with CO_2 limitation.

Figure 4.1 15% SDS-PAGE of lysates of *Sulfolobus* LM cells grown under air and 5% (v/v) CO_2 in air, including low molecular weight standards (LMW Stds). A band occurring at approximately 19 kDa in the track containing air-grown cells but not in the track containing cells grown using 5% (v/v) CO_2 in air is arrowed.

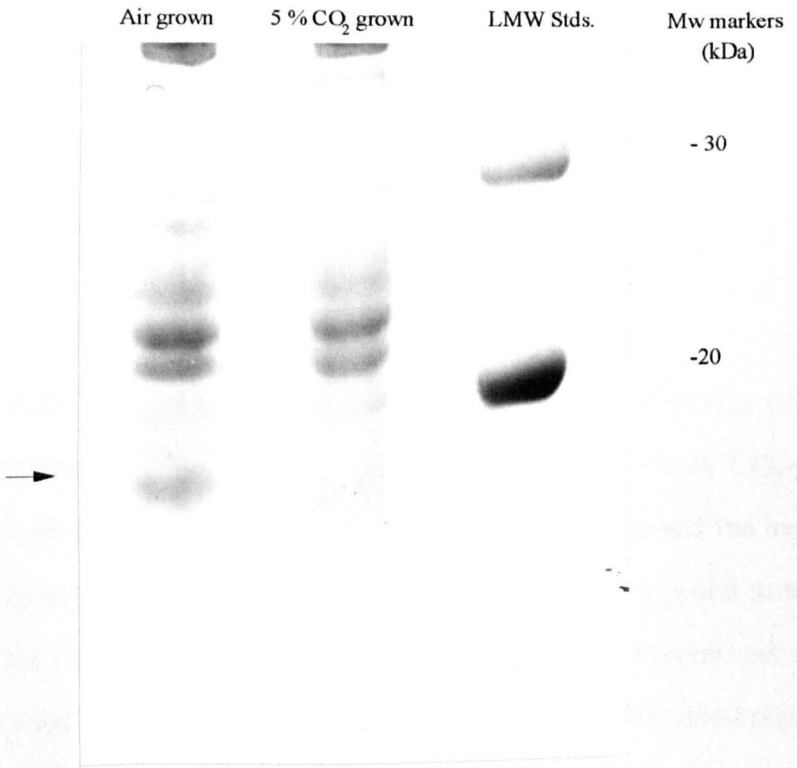


Figure 4.2 10% SDS-PAGE of cell lysates of *Sulfolobus* LM grown under air (AIR) and 1% CO_2 in air (1% CO_2) including low molecular weight standards (LMW std). A band at approximately 59 kDa is arrowed. This is present in excess in CO_2 limited cells.

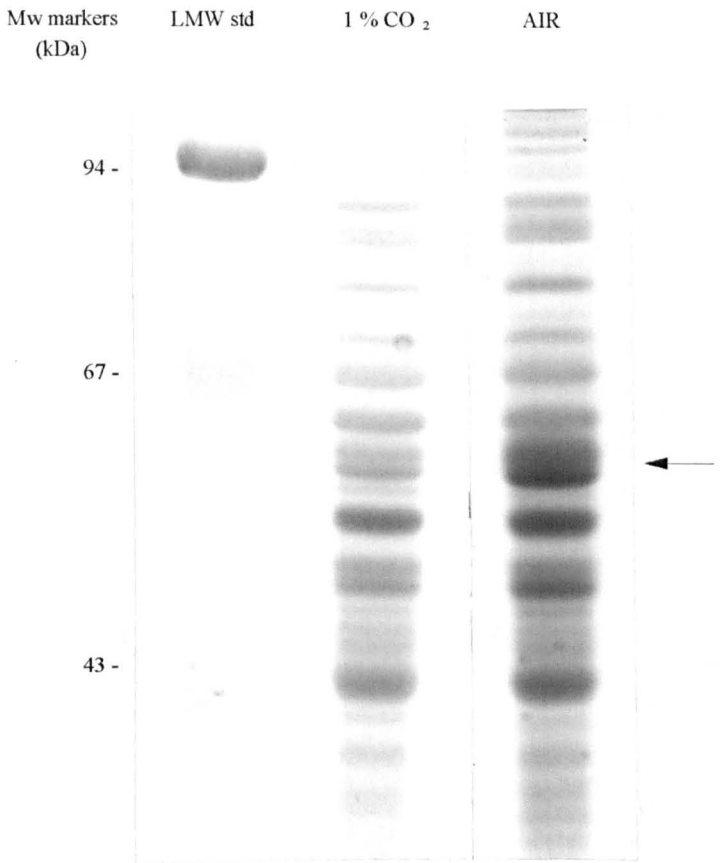
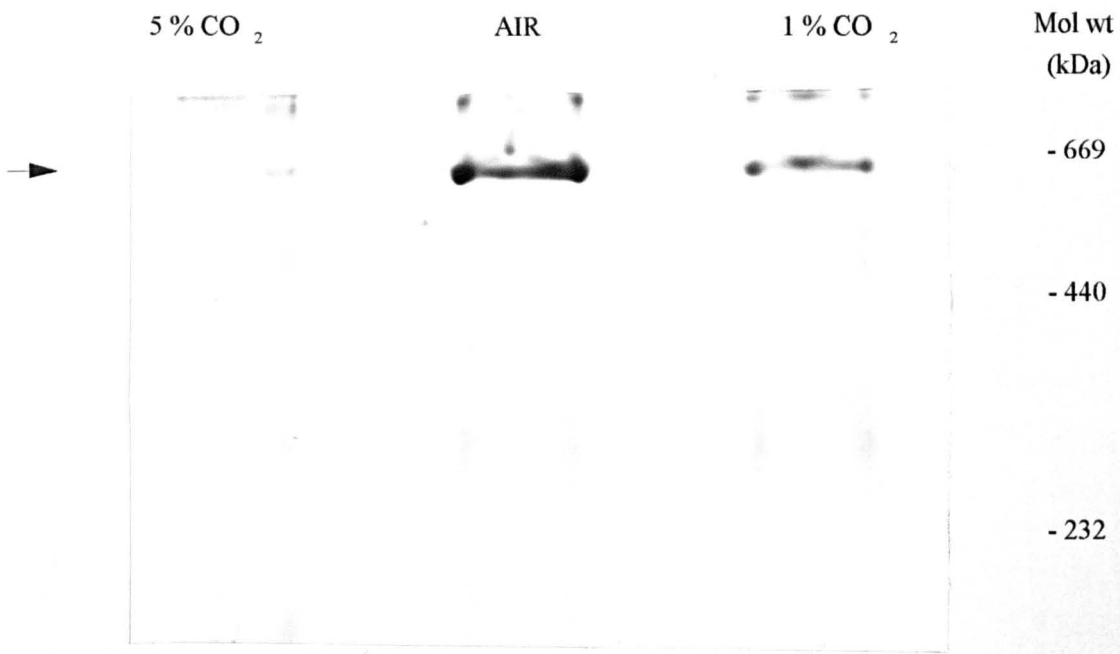


Figure 4.1 clearly showed the increased amount of one protein band in the sample prepared from air-grown cells as opposed that from CO_2-grown cells. This band ran at approximately 19 kDa. Figure 4.2 clearly showed the increased amount of another protein band in the sample of air-grown cells as opposed to that of CO_2-grown cells. This band ran at approximately 59 kDa. These observations were reproducible over five different gels, and agreed with the previously published results of Norris *et al* ,(1989).

4.3 - Native PAGE of proteins from cells grown with CO_2 limitation and CO_2 excess

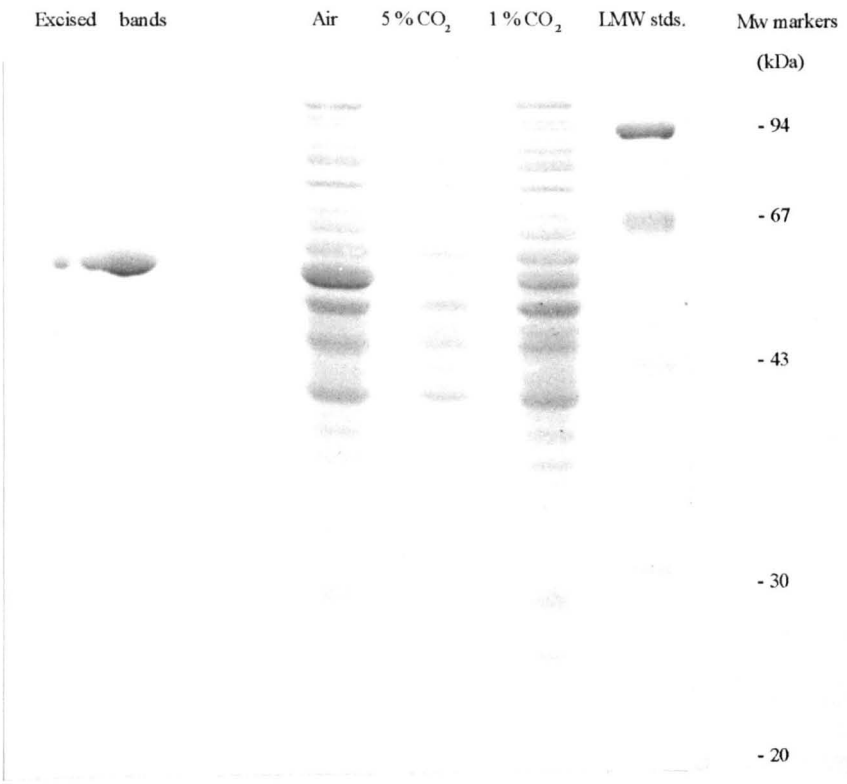
Native PAGE was carried out using samples of *Sulfolobus* LM as above. Gels were examined for the presence of increased proportions of protein bands in the samples grown under air.

Figure 4.3 Native PAGE of cell lysates of *Sulfolobus* LM grown under air (AIR), 5% (v/v) CO_2 in air (5% CO_2) and 1% (v/v) CO_2 in air (1% CO_2).



The native polyacrylamide gel showed one protein band increased in relative concentration during growth under limiting CO_2-conditions. This appeared to be one of the most abundantly produced proteins of autotrophically-grown *Sulfolobus* LM. It ran at a position slightly below the 669 kDa marker on a native gel. The actual molecular weight of this native protein was undetermined. The excision of this protein band from the native gel was followed by SDS sample treatment and SDS-PAGE.

Figure 4.4 SDS-PAGE of the air-growth induced band excised from a native polyacrylamide gel (Excised bands). Lysate from CO_2-limited cells (Air), cells grown under CO_2 excess (5% CO_2 and 1% CO_2) and low molecular weight standards (LMW stds.) were electrophoresed.



This SDS polyacrylamide gel (Fig. 4.4) showed that the native protein giving a band below the 669 kDa marker (Fig. 4.3) was made up of 59 kDa and 19 kDa subunits. These corresponded to those seen by single step SDS-PAGE (Figs. 4.1 and 4.2) and with those identified by Norris *et al,* (1989). Therefore the confirmation of these previous results was achieved. This showed that the protein described by Norris *et al,* (1989) was indeed the one upon which further investigation was based.

4.5 - N-terminal amino acid sequencing of 59 kDa subunit

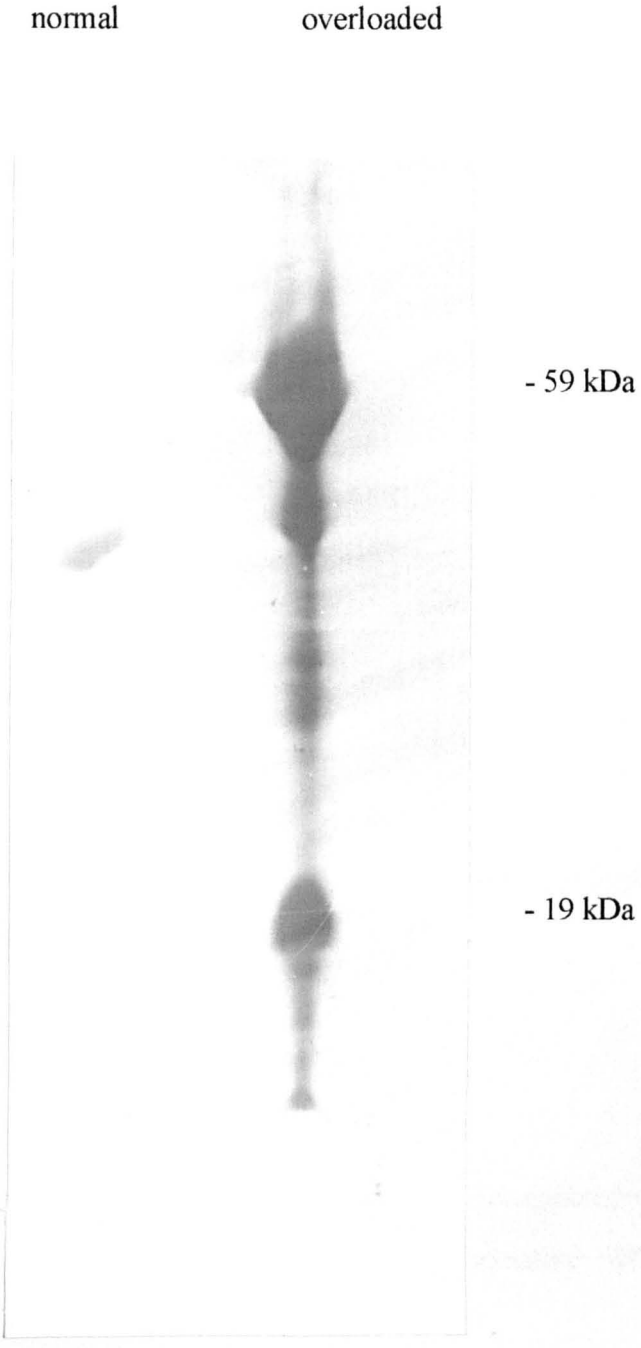
The partial purification of small quantities of both the 59 kDa and the 19 kDa subunits was achieved by polyacrylamide gel electrophoresis. This required the growth of gram quantities of *Sulfolobus* LM under limiting CO_2 conditions, and therefore the use of a 20 l fermenter sparged with air only. Native polyacrylamide gel electrophoresis and Coomassie staining was carried out on 500 mg of these cells, yielding a large band of protein at a position corresponding to the 600 kDa marker. This was immediately excised and subjected to the SDS sample preparation procedure, then used for SDS-PAGE. One track was run of normal protein concentration, another was grossly overloaded in an attempt to concentrate the maximum quantity of material for N-terminal sequencing. The SDS gel was subsequently blotted onto PVDF membrane which was then Coomassie stained and photographed (Fig. 4.6 overleaf). However, this procedure resulted in poor picture quality.

Although the purification was by no means complete, it was estimated that both the 59 kDa and the 19 kDa band were present at many times the concentration of the third most abundant protein in the preparation. The area of membrane containing the 59 kDa band was sent for N-terminal amino acid microsequencing. Sixteen amino acids were identified from the N-terminus of this polypeptide (Fig. 4.5). However, the final four amino acids were reported to be uncertain.

Figure 4.5 N-terminal amino acid sequence of 59 kDa polypeptide

NH_2-Pro-Pro-Phe-Gly-Lys-Val-Leu-Val-Ser-Asn-Arg-Gly-Glu-Ile-Ala-Val-

Figure 4.6 PVDF membrane blot of SDS-PAGE of the excised native protein band overproduced by CO_2 limited cells of *Sulfolobus* LM. Tracks contain normal (normal) and overloaded (overloaded) samples of the denatured protein. The 59 kDa and 19 kDa bands are indicated.



4.6 - Design of DNA probe to the gene for the 59 kDa protein

The N-terminal sequence data was back translated using the genetic code and represented as follows ;-

NH_2-Pro-Pro-Phe-Gly-Lys-Val-Leu-Val-Ser-Asn-Arg-Gly-Glu-Ile-Ala-Val-

```
CCG CCG TTT GGG AAG GTG TTG GTG AGT AAT CGG GGG GAG ATA GCG GTG
 A   A   C   A   A   A   A   A   C   C   A   A   A   T   A   A
 T   T       T       T C G   T TCG       T   T       C   T   T
 C   C       C       C C A   C   A       C   C           C   C
                       C T       T
                       C C       C
```

As the C-terminal portion of this sequence was uncertain, the N-terminal portion of the sequence was chosen for probe production. This portion avoided the wobble positions of leucine and serine which would have necessitated increasing the degeneracy of the DNA probe greatly. A 20-mer oligonucleotide, suitable for probing and sequencing, was designed using the codon bias of *Sulfolobus acidocaldarius* thermopsin (Lin *et al.*, 1990), *Sulfolobus acidocaldarius* sox terminal oxidase complex (Lübben et al., 1992), and *Sulfolobus solfataricus* aspartate aminotransferase (Cubellis *et al.*, 1989). These sequences all show a bias towards A and T in the wobble positions of proline, lysine and valine, as would be expected in these organisms of low GC content.

The antisense probe sequence was constructed, therefore, the sequence was as follows ;-

```
5'-AG TAC TTT TCC AAA TGG TGG-3'
   A A   C   A   G   A   A
```

This oligonucleotide probe was therefore constructed with 128 degeneracies, it had a theoretical melting temperature of approximately 53°C. This oligonucleotide was designated 59-1.

4.7 - Detection of the gene coding for the 59 kDa protein in *Sulfolobus* LM

Restriction digests of genomic DNA from *Sulfolobus* LM using *Hin*d III were Southern blotted and probed using DIG labelled probe 59-1. Hybridisations were carried out at 37°C, 40°C and 45°C with 5 x SSC and 2 x SSC washes. Bound oligonucleotide was detected luminescently (Fig. 4.7).

Figure 4.7 Autoluminescent photograph of a Southern blot of *Hin*d III digested *Sulfolobus* LM genomic DNA probed with oligonucleotide 59-1 at 40°C. The major hybridised band is indicated.

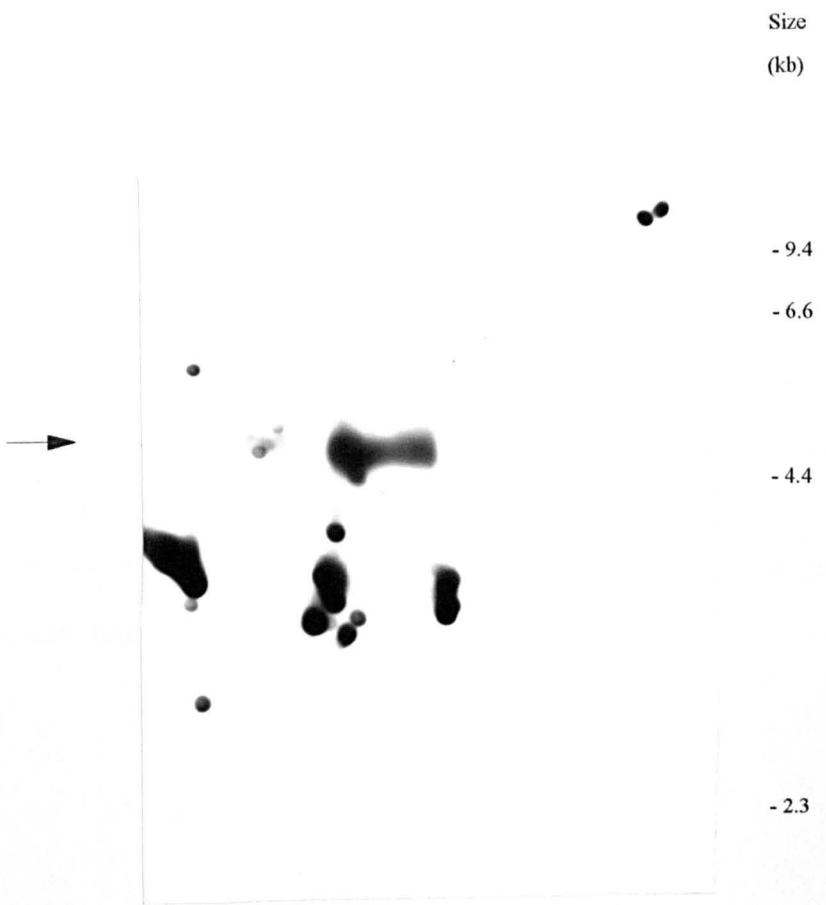


Figure 4.7 showed a major band of oligonucleotide binding, which by reference back to standard DNA run on the original agarose gel can be said to have hybridised to a genomic DNA fragment of approximately 5.2 kb. This result was confirmed by two additional similar experiments.

4.8 - Isolation of a clone homologous to oligonucleotide 59-1

A partial library of *Sulfolobus* LM genomic DNA fragments of between 6 and 4.5 kb was constructed. The fragments were formed by restriction with *Hin*d III then eluted from an agarose gel. A pBluescript KS+ phagemid vector was chosen as the cloning vector. This is a high copy number vector which has a large polylinker and ampicillin resistance and is amenable to blue / white selection of transformants with inserts. The vector was restricted with *Hin*d III, phosphatased to prevent self religation and repurified. After purification, the genomic DNA fragments were ligated into the vector, then transformed into *E. coli* strain TG 1. 500 white colonies were selected and replated in duplicate. One set of colonies was colony blotted and hybridised with probe 59-1 at 45^{o}C with washes in 5x and 2x SSC at the same temperature (Fig. 4.8)

Figure 4.8 Autoluminescent photograph of a colony blot of a partial *Sulfolobus* LM genomic library of *Hin*d III fragments between 6 and 4.5 kb, probed with oligonucleotide 59-1.

Five possible positive colonies were highlighted by hybridisation to the probe. These were all picked, cultured and used for plasmid preparations. The plasmid DNA was used for a dot blot hybridisation with probe 59-1 utilising pBluescript as a negative control. This gave positive hybridisation for the transformed plasmids, but no hybridisation using solely pBluescript. This confirmed the colony blot result, signifying the isolation of clones containing DNA sequence homologous to that of the oligonucleotide probe 59-1.

4.9 - Initial characterisation of positive clones

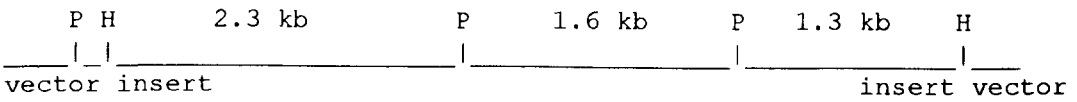
Plasmid DNA of all five positive clones was doubly digested using *Hin*d III and *Eco* RV, *Hin*d III and *Pst* I and *Hin*d III and *Xho* I. All clones gave similar results with the insert being excised in all cases, but further digestion of the insert occurring only with *Pst* I. All clones gave identical restriction patterns using *Pst* I plus *Hin*d III, giving fragments of approximately 2.3, 1.6 and 1.3 kb. This showed that all clones were of identical fragments although their orientation in the polylinker may have been different. Southern blots of the digested plasmids hybridised with probe 59-1 at 45°C showed probe binding only to the 2.3 kb fragment.

4.10 - Subcloning of the 2.3 kb fragment

A 59 kDa protein would be expected to require a gene of around 1.5 kb. Therefore some subcloning was carried out in order to reduce the amount of DNA sequencing required. *Pst* I digestion of one clone, 5.2-1, resulted in the release of the 2.3 and 1.6 kb bands from the backbone. Therefore the 1.3 kb fragment must have been situated at the opposite end of the insert from the vector polylinker *Pst* I site. This meant that the 2.3 kb fragment may have had two *Pst* I termini or one *Hind* III and one *Pst* I terminus.

The 5.2-1 clone was doubly digested with *Hind* III and *Pst* I and shotgun cloned into pBluescript restricted with *Pst* I only, and also into pBluescript restricted with both *Hind* III and *Pst* I. The *Hind* III / *Pst* I ligation produced, amongst others, some clones containing solely the 2.3 fragment whereas the *Pst* I ligation produced no clones containing solely the 2.3 kb fragment. This necessitated the following structure for the 5.2-1 clone (Fig. 4.9).

Figure 4.9 Restriction map of clone 5.2-1



P - *Pst* I restriction site
H - *Hind* III restriction site

N.B. probe 59-1 binds only to the 2.3 kb fragment

The clone containing solely the 2.3 kb fragment was digested with *Hind* III and *Pst* I, producing the vector backbone and the 2.3 kb fragment which hybridised to probe 59-1 in a further Southern blot. This construct was designated p2.3.

4.11 - Confirming that p2.3 codes for the 59 kDa protein

Although a 2.3 kb fragment of DNA to which probe 59-1 was homologous had been cloned, it remained to be shown that this represented the gene coding for the 59 kDa protein. This was achieved by sequencing downstream from the region to which the probe 59-1 bound. If this region of DNA translated to give an amino acid sequence identical to that derived from N-terminal sequencing of the 59 kDa protein this would show that the gene coding for this protein had been isolated.

Manual plasmid sequencing was carried out using p2.3 as template with oligonucleotide 59-1 and M13 reverse primer, homologous to a site on pBluescript adjacent to the polylinker, being used as sequencing primers. Fortuitously, the priming site for 59-1 was found to be only 36 nucleotides away from the *Hin*d III site with which the entire fragment was cloned. This meant that sequence from the M13 reverse and 59-1 primers overlapped, with some sequence downstream from the 59-1 priming site being revealed. The revealed sequence translated into an identical amino acid sequence to that derived from N-terminal sequencing of the 59 kDa protein (Fig. 4.10).

Figure 4.10 Initial DNA sequence of clone p2.3 primed with M13 reverse primer compared with pBluescript sequence. Also amino acid translation of clone p2.3 compared with the N-terminal sequence of the 59 kDa protein.

```
pBluescript -     aaccctcactaaagggaacaaaagctgggtaccgggccccccctcgaggt
p2.3        -     aaccctcactaangggaacaaaagctggntaccgggccccccctcgaggt

                            HindIII
pBluescript -     cgacggtatcgataagctt
p2.3        -     cgacggtatcgataagctttttttaaactatactaaacattataagttgag


p2.3        -     atgccaccatttggaaaagtacttgtttcaaatagaggagagatagccgta
59-1 site   -        ^^^^^^^^^^^^^^^^^^^^^^
translated  -     M  P  P  F  G  K  V  L  V  S  N  R  G  E  I  A  V
N-terminal  -        P  P  F  G  K  V  L  V  S  N  R  G  E  I  A  V
```

4.12 - Restriction mapping of the 2.3 kb fragment

To facilitate the sequencing of the 2.3 kb fragment, further subcloning of the fragment was to be carried out. This required a restriction map of the fragment. Digests were carried out on p2.3 using restriction enzymes with sites present in the pBluescript polylinker and also enzymes with no sites or few sites in the pBluescript backbone but possessing ends compatible with sites present in the polylinker.

The following restriction enzymes were used and found not to cut within the fragment ;-

Bam HI	*Kpn* I	*Sac* I	*Sal* I	*Spe* I
Xba I	*Xho* I	*Pst* I	*Hind* III	*Eco* RV
Sma I	*Not* I	*Bgl* I	*Hinc* II	*Nar* I
Hpa I	*Nru* I			

Six cutter restriction enzymes tested which did cut within the fragment were ;-

Eco RI	*Bgl* II	*Acc* I

Four cutter restriction enzymes tried which did cut within the fragment were ;-

Sau 3AI	*Alu* I	*Hae* III	*Taq* I

The most useful of these restriction enzymes in terms of suitability for subcloning were *Eco* RI and *Bgl* II which would produce fragments which could be ligated directly into, respectively, the *Eco* RI and *Bam* HI sites of the pBluescript polylinker. *Acc* I was potentially useful as an *Acc* I site is present in the pBluescript polylinker but flawed by the lack of specificity of the enzyme, which can cut at four different sequences only one of which could be cloned into the pBluescript polylinker.

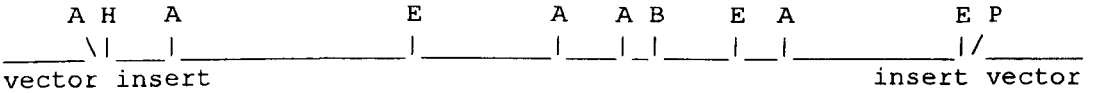
Eco RI digestion of p2.3 produced fragments of approximately 800 bp and 600 bp. When doubly digested with *Hind* III an additional fragment of 850 bp was produced. When doubly digested with *Pst* I no additional fragment was visualised.

Bgl II digestion of p2.3 caused linearisation of the plasmid, showing that only one *Bgl* II site is present in the insert. When doubly digested with *Hind* III a 1500 bp fragment was produced. When doubly digested with *Pst* I an 800 bp fragment was produced.

There was one *Acc* I site present in the pBluescript polylinker of p2.3, this was situated close to the *Hind* III site. *Acc* I digestion of p2.3 produced fragments of 150 bp, 200 bp, 1050 bp and 350 bp. Double digestion using *Pst* I produced one extra fragment of 500 bp.

Acc I digestion of the 600 bp *Eco* RI fragment gave one 500 bp fragment and one fragment of just over 100 bp. *Acc* I digestion of *Eco* RI digested backbone produced fragments of 150 bp and 700 bp. *Acc* I digestion of the 800 bp *Eco* RI fragment gave fragments of 350 bp and a doublet at 250 bp. These observations allowed a restriction map to be deduced (Fig. 4.11).

Figure 4.11 Restriction map of p2.3 as determined by single and double digestions. Total length 2.3 kb from *Hind* III to *Pst* I site, drawn to scale.



A - *Acc* I restriction site
H - *Hind* III restriction site
E - *Eco* RI restriction site
B - *Bgl* II restriction site

N.B. The 59-1 binding site is 36 bases downstream from the Hind III site

4.13 - Further subcloning from p2.3

The primary objective of these experiments was to facilitate DNA sequencing by subcloning the three fragments produced by *Eco* RI digestion of p2.3 and by subcloning the two fragments produced by *Bgl* II digestion of p2.3.

No ligations of fragments made by *Acc* I digestion were successful, suggesting that the *Acc* I sites found within p2.3 were not compatible with those of the pBluescript polylinker.

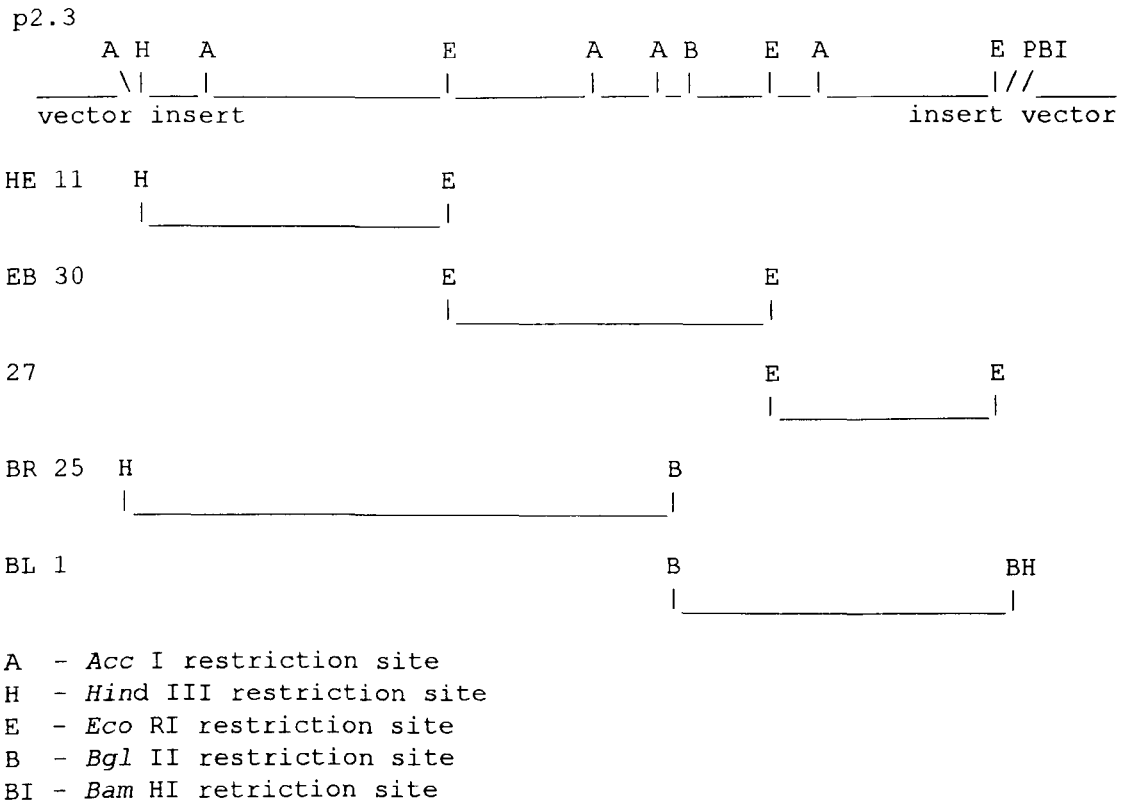
The 3' *Eco* RI - *Eco* RI fragment of p2.3 was subcloned simply by digesting p2.3 with *Eco* RI and shotgun cloning the resultant mix of fragments into *Eco* RI digested pBluescript. The clone containing a single 600 base *Eco* RI - *Eco* RI fragment was digested with *Acc* I and found to be cut into one 500 base and one 120 base fragments. This was designated clone 27.

The first 850 and the second 800 bases of p2.3 were subcloned by digesting p2.3 with *Hin*d III and *Eco* RI, thus releasing an 850 base *Hin*d III - *Eco* RI fragment, an 800 base *Eco* RI - *Eco* RI fragment and the 600 base fragment subcloned previously. The resulting mix of fragments was digested with *Bam* HI, an enzyme on the polylinker of pBluescript 3' to the *Pst* I site. This prevented religation of the original plasmid backbone with any of the fragments by removing the *Eco* RI site. This mixture was purified then ligated into fresh pBluescript restricted using *Hin*d III plus *Eco* RI and also ligated into fresh pBluescript restricted using only *Eco* RI. The first, 850 base *Hin*d III - *Eco* RI fragment was isolated in a clone designated HE 11. The second, 800 base *Eco* RI fragment was isolated in a clone designated EB 30.

The subcloning of the *Bgl* II fragments was facilitated by the compatibility of overhangs produced by *Bgl* II with those produced by *Bam* HI. The 3' 800 base *Bgl* II

fragment was subcloned by digesting p2.3 with *Bgl* II and *Bam* HI, purifying the 800 base fragment using Qia-ex treatment then ligating it into fresh pBluescript restricted with *Bam* HI. The successful clone was designated BL 1. The 5' 1500 base fragment was subcloned by digesting p2.3 with *Bgl* II and *Bam* HI then simply religating. Clones containing solely the 1500 base fragment were picked with one being designated BR 25.

Figure 4.12 Subclones of p2.3 in relation to the original clone. Drawn to scale.



These subclones (Fig. 4.12) and p2.3 were sequenced manually and automatically from both ends using M13 reverse primer and M13 -20 primer, sites to both of which exist just outside the pBluescript polylinker. Some regions of sequence, where overlapping sequence from both strands of DNA was not available, required

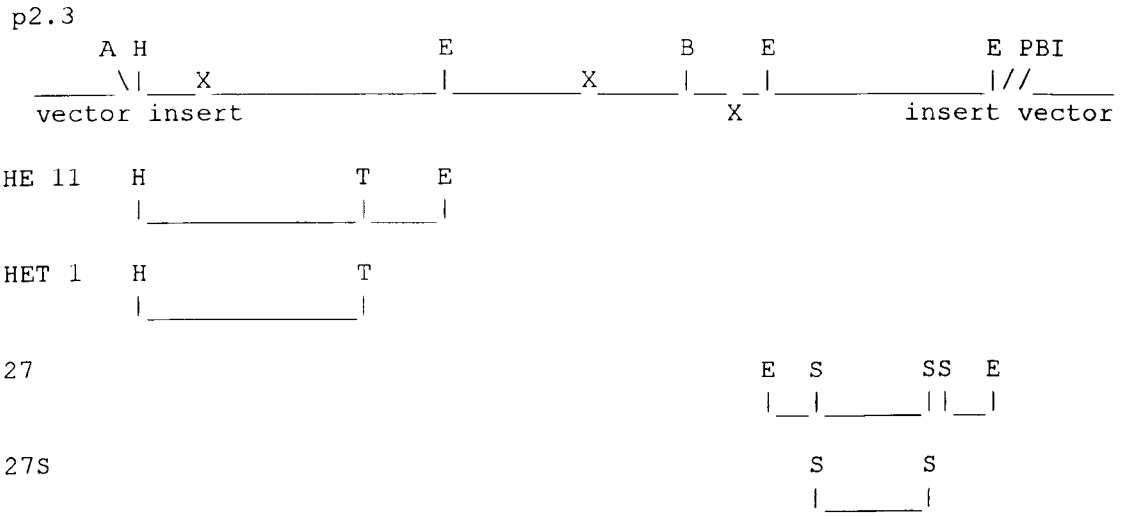
further investigation. This was provided by the construction of two additional subclones and three internal primers.

The subclone HE 11 was further subcloned using *Taq* I restriction enzyme, which is compatible with *Cla* I overhangs. A *Taq* I site was situated approximately 650 bases away from the *Hin*d III end of HE 11. HE 11 was digested with *Hin*d III and *Eco* RI, separated by electrophoresis and the 850 base *Hin*d III - *Eco* RI fragment purified by the Qia-ex procedure. This was then digested using *Taq* I, which requires a high temperature for full activity. The resulting mixture was ligated into pBluescript previously cut using *Hin*d III and *Cla* I. The resulting construct, named HET 1 was sequenced automatically using M13 primers.

The subclone 27 was further subcloned using *Sau* 3AI restriction enzyme, which is compatible with *Bam* HI overhangs. *Sau* 3AI sites were situated at approximately 100, 500 and 520 bases away from the 5' end of clone 27. Clone 27 was digested with *Eco* RI , separated by electrophoresis and the 650 base *Eco* RI - *Eco* RI fragment purified by the Qia-ex procedure. This was then digested using *Sau* 3AI. The resulting mixture was ligated into pBluescript previously restricted using *Bam* HI. A transformant possessing the 400 base insert was selected. The resulting construct, named 27S was sequenced automatically using M13 primers.

Additional 20 mer sequencing primers were designed for positions 100 and 1300 to sequence the top strand, and position 1600 to sequence the reverse strand. The 100f primer was used to sequence subclone HE 11. The 1300f and 1600r primers were used to sequence subclone EB 30.

Figure 4.13 Additional subclones of p2.3 in relation to the original clone. Drawn to scale.

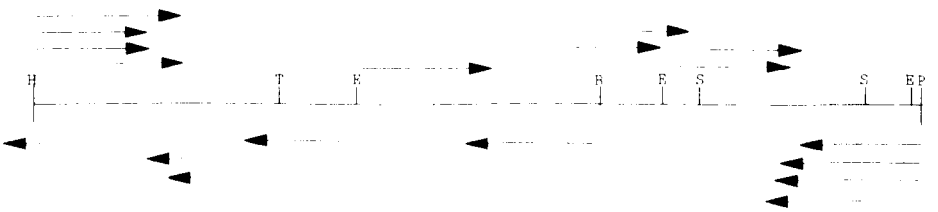


A - *Acc* I restriction site
H - *Hind* III restriction site
E - *Eco* RI restriction site
B - *Bgl* II restriction site
BI - *Bam* HI restriction site
T - *Taq* I restriction site
S - *Sau* 3AI restriction site
P - *Pst* I restriction site
X - Binding sites of new primers

4.14 - Sequencing of clone p2.3

Plasmid DNA sequencing was carried out using both manual and automatic sequencing techniques. Ideally, the entire sequence of both DNA strands would have been elucidated. However, time was only available to sequence sections of both strands which could be assembled into the primary DNA sequence of the whole fragment (Fig. 4.14). This may well have lead to minor sequence inaccuracies.

Figure 4.14 Diagram of sequencing experiments carried out upon p2.3 and its component subclones. This represents the entire length of p2.3, with restriction sites as shown in figure 4.13 above. Arrows represent sequences obtained and the direction of sequencing. Drawn to scale.



4.15 - DNA and protein sequence

The DNA sequence derived from sequencing p2.3 (see Fig. 4.15) was translated into all three forward frames. The N-terminal sequence data had shown the correct frame for the 59 kDa protein. The amino acid immediately preceeding this sequence was found to be methionine, produced by the codon ATG which represents the translation start codon. This must represent the start codon for the 59 kDa protein, as a stop codon (TAA) is present only nine bases upstream. Post translational cleavage of the methionine from this protein is likely to have occurred. The case for recognising this sequence as the start codon of the 59 kDa protein is strengthened by the presence immediately upstream from the ATG codon of a possible ribosome binding site and possible transcriptional promoter regions homologous to box A and box B (see section 4.16). Therefore for further analysis of this protein the start codon was assumed to be the ATG codon at position 37 of the DNA sequence of p2.3.

Reading through the translation of this sequence it was apparent that the open reading frame continued for a further 512 amino acids until two stop codons (TAATGA) were encountered. Immediately downstream of this position was found another open reading frame continuing for around 157 amino acids. This was found on a different frame from the first polypeptide but no N-terminal amino acid sequence data was available for a protein produced by this open reading frame. Therefore its start codon was more difficult to assign with a great degree of confidence. Possible start codons (ATG) were present 51 bases upstream of, 3 bases downstream from, 24 bases downstream from and 87 bases downstream from the end of the previous polypeptide. The most likely actual start codon was considered to be the one found 3 bases downstream from the end of the previous polypeptide due to its position and the presence 8 bases upstream of a possible ribosome binding site, an advantage not shared by any of the other candidates. For further analysis this was assumed to be the actual start codon for the second open reading frame.

4.16 - DNA structure

The DNA sequence derived from p2.3 showed, apart from the open reading frames, some recognised consensus sequences of archaeal DNA structure. These included transcriptional promoters similar to box A and box B, possible ribosome binding sites and transcriptional terminator sequences.

The sequence similar to transcriptional promoter box A is situated between 36 and 25 bases upstream from the translational start codon. This sequence (AAGCTTTTTAAA) is identical to box A of the tRNA methionine gene of *Thermoproteus tenax* (Brown *et al.*, 1989) and conforms to the consensus sequence TTTA (A or T) A (Reiter *et al.*, 1990).

The sequence similar to box B is situated between 2 and 5 bases upstream from the start codon. This possesses the sequence TTGA, conforming to the consensus sequence (A or T) TG (A or C) (Reiter *et al.*, 1990).

The first putative ribosome binding site is located actually at the translational start codon, next to the putative box B. This sequence is AGATG, similar to the *Sulfolobus acidocaldarius* consensus ribosome binding site (A or G) GGTG (Olsen *et al.*, 1985). This sequence is usually found 6-8 bases upstream from the initiation codon, so it would appear that either the translation apparatus is extremely compact or this does not represent the ribosome binding site. If the second case is true then the ribosome binding site may be present on a stem loop structure further downstream as for the *bop* gene of *Halobacterium halobium* (Brown *et al.*, 1989). No structure of this type has yet been identified. More definitive information on this subject could be provided by studies of the mRNA transcript.

The second putative ribosome binding site is present between 10 and 6 bases upstream of the initiation codon of the second open reading frame. This is a more usual position for such a site, and it possesses the same sequence as the first site (AGATG). As this sequence is actually present within the coding area of the first cistron, this supports the proposed co-transcription of these two cistrons.

Probable transcription termination signals of TTTTTT and TTTTT are present downstream of the stop codon of the second cistron. These sequences are identical to those of Sulfolobus VLP SSV1 genes (Brown *et al.*, 1989).

Figure 4.15 DNA sequence of clone p2.3 with translation of the 59 kDa protein and the subsequent open reading frame.

```
                'box A'                          'box B'RBS
  1  AAGCTTTTTAAACTATACTAAACATTATAAGTTGAGATGCCACCATTTGGAAAAGTACTT
                                                  *****
  1                                                M  P  P  F  G  K  V  L

 61  GTTTCAAATAGAGGAGAGATAGCCGTAAGCGTAATGAAGAGAATTAAAGAGATGGGAATG

  9  V  S  N  R  G  E  I  A  V  S  V  M  K  R  I  K  E  M  G  M

121  AAGGCAGTAGCTGTCTACTCTGAGGCAGATAAATATGCCCTCCACGTAAAATATGCGGAC

 29  K  A  V  A  V  Y  S  E  A  D  K  Y  A  L  H  V  K  Y  A  D

181  GAAGCCTACTACATAGGACCTGCACCTTCTATTGAGAGTTATCTGAACATTGAAAGGATT

 49  E  A  Y  Y  I  G  P  A  P  S  I  E  S  Y  L  N  I  E  R  I

241  ATAGATGCAGCGGAGAAAGCACATGCAGACGCAGTACATCCGGGTTACGGTTTCCTGTCG

 69  I  D  A  A  E  K  A  H  A  D  A  V  H  P  G  Y  G  F  L  S

301  GAGAGAGCAGACTTCGCTGAGGCCGTAGAGAAGGCAGGTATGACTTTCATTGGACCCTCA

 89  E  R  A  D  F  A  E  A  V  E  K  A  G  M  T  F  I  G  P  S

361  TCTGAGGTAATGAACAGGGTAAAGAGTAAGCTAGACGGAAAAGAATTGCAAAGAAGCAGG

109  S  E  V  M  N  R  V  K  S  K  L  D  G  K  E  L  Q  R  S  R

421  AACGCCTATTCCCCCGGATCTGACGGACCAGTTGGATCGTTGGATGAGGCATTGAAACTT

129  N  A  Y  S  P  G  S  D  G  P  V  G  S  L  D  E  A  L  K  L

481  ACAGAGAAAATAGGATACCCAGTAATGGTAAAAGCAGCTTATGGAGGTGGAGGTATAGGG

149  T  E  K  I  G  Y  P  V  M  V  K  A  A  Y  G  G  G  G  I  G

541  ATAACAAGAGCCGACAATCCTGATCAGTTGATGGACATATGGGGTAGAAACCAAAGGTTA

169  I  T  R  A  D  N  P  D  Q  L  M  D  I  W  G  R  N  Q  R  L

601  GCTAAGGAAGCCTTCGGTAGGCCAGACTTGTATATAGAGAAAGCTGCGGTTAGACCGAGA

189  A  K  E  A  F  G  R  P  D  L  Y  I  E  K  A  A  V  R  P  R

661  CACATCGAGACTCAGCTTATTGGAGATAAATACGGAACTTATGTAGTTGCATTTGAAAGG

209  H  I  E  T  Q  L  I  G  D  K  Y  G  T  Y  V  V  A  F  E  R
```

721 GAGTGTCACAATCAGAGGAGAAATCAGAAACTTATAGAGGAGGCTCCGTCTCCGGCCCTA

229 E C H N Q R R N Q K L I E E A P S P A L

781 ACTTGGGAAAAGAGGGAAGAGTTCATTGATGCATCAGTTAAGTACGGCAAGAAGATTGGT

249 T W E K R E E F I D A S V K Y G K K I G

841 TACTTCGCTTTAGGAACTATGGAATTCGCATACTCTGACACGACCAAGGATCTGTACTTC

269 Y F A L G T M E F A Y S D T T K D L Y F

901 CTGGAGCTTAACAAGAGATTACAAGTAGAGCATGGAATTACTGAGCTTGTAACTGGGATC

289 L E L N K R L Q V E H G I T E L V T G I

961 GATCTAGTGAAACTCCAAATAAGGTTAGTAGCAGGAGAACATATGCCTTTCACTCAGGAG

309 D L V K L Q I R L V A G E H M P F T Q E

1021 GAACTAAAGAAGAGACTCAGAGGACACGCAATAGAATATAGGATAAACGCTGAAGATCCG

329 E L K K R L R G H A I E Y R I N A E D P

1081 TTAAACAACTTCACCGGTAGTTCAGGTTTTGTAACATATTATAAGGAACCAAGCGGACCT

349 L N N F T G S S G F V T Y Y K E P S G P

1141 GGAGTGAGGCTTGATACAGGGATAATGGAAGGAAGGTATGTACCTCCATTCTATGATTCA

369 G V R L D T G I M E G R Y V P P F Y D S

1201 TTGGTAGCTAACCTCATGGTATACGGAGAGAACAGATCAGGGGCGATAGAAGTAGGAAGG

389 L V A N L M V Y G E N R S G A I E V G R

1261 AGATCATTGAGAGACCTACAGATAGGAGGTATAAAGACCACAATACCTCTCTACAAGCTG

409 R S L R D L Q I G G I K T T I P L Y K L

1321 ATAATGGATGACGAAGATTACCAGAATGGAAACTTCACTACCGCATACATAGCGGATAAG

429 I M D D E D Y Q N G N F T T A Y I A D K

1381 AGTGAGATATTCACAAAGAAGCTCAGGGAGAAGGAAATGCTGAAGACAGCTTTAGCTGCA

449 S E I F T K K L R E K E M L K T A L A A

1441 TCAGTCTACAATAGAGGATTGCTGAGAGGATCTAGCGGAAATAACAACGTACAAGCTCAA

469 S V Y N R G L L R G S S G N N N V Q A Q

```
1501    GATAACGGCAAACCTAGATCTGCATGGAAGACATATGGAGTTCAGGCTCAGGCTCCTGCA
 489    D  N  G  K  P  R  S  A  W  K  T  Y  G  V  Q  A  Q  A  P  A

             RBS
1561    AGGGTGATGTGGTAATGAAGCTCCTCAGAGTTTACATGGAAACTGGAGAAACCTACATTG
          *****
 509    R  V  M  W  *  *

   1                  M  K  L  L  R  V  Y  M  E  T  G  E  T  Y  I  A


1621    CAAGCTATGACCAGAAGGATAACAAGGACACAGTGAAAATGGAAGAAGGAGAGTTAAATG
  17      S  Y  D  Q  K  D  N  K  D  T  V  K  M  E  E  G  E  L  N  V

1681    TAGAATTCTTAGGTAGAGGTACAAGAGAAAACGAGTACTTGTTCAAGGTTGGAAACGAAG
  37      E  F  L  G  R  G  T  R  E  N  E  Y  L  F  K  V  G  N  E  V

1741    TACACAGCATTACAATTGATAGGGGATTCCTAATATTAGATCAGGAAGAAGAGTATAAGG
  57      H  S  I  T  I  D  R  G  F  L  I  L  D  Q  E  E  E  Y  K

1801    TAGACAGGATAACAGAACTTCCAGTAAAGGAGGGTCAATCAGTAGAGGAATTAATGAAAG
  77      D  R  I  T  E  L  P  V  K  E  G  Q  S  V  E  E  L  M  K  G

1861    GAAAAGAAGGAGAAGTATTATCGCCATTGCAAGGTAGAGTTGTAGCTATAAGAGTAAAGG
  97      K  E  G  E  V  L  S  P  L  Q  G  R  V  V  A  I  R  V  K  E

1921    AAGGTGACGCAGTAACTAAAGGTCAGCCTCTTCTATCTGTAGAAGCAATGAAATCTGAGA
 117      G  D  A  V  T  K  G  Q  P  L  L  S  V  E  A  M  K  S  E  T

1981    CCATAATATCTGCACCAATAGCTGGAGTTATAGAGAAAATTGCGGTAAAAACCAGGCCAG
 137      I  I  S  A  P  I  A  G  V  I  E  K  I  A  V  K  T  R  P  G

2041    GGAGTAAAGAAGAGGAGACTTGCTGGTTGTACTAAAATAAATTCTGAAAACGATTGTTTC
 157      S  K  E  E  E  T  C  W  L  Y  *

                term ?                                  term ?
2101    TATAATAGATTGATTTTTTGCTATAGATATTATATTTATATATTTTTTATATTATTTATT
                ^^^^^^                                  ^^^^^^
                                                 term ?
2161    TTATTACTATTTCTACTATCCAATCTCTTCTCATGTATTTTTCTTAATGCTCTAGTTGAT
                                                 ^^^^^
2221    CCAGTTAAAGATCCACGAGAAGTAACGTCAGAACAGATAGTCTAAAGCTGATAAAAATTT
2281    ACATACTATTTGTCTCATTTAACTCTTATGAACATGGAGAAAGGTCTTTCATTGGAAAAA
2341    GAATTCGTCGTGGAAAATCAGCATTCTGCAGC
```

4.17 - Identification of protein homologies

Totalling the molecular weights of the amino acids of the two polypeptides produced by these two open reading frames gave molecular weights of 56,980 Da and 17,787 Da. These are slightly less than the 59 kDa and 19 kDa molecular weights of the two polypeptides observed by SDS-PAGE. The 59 kDa polypeptide had been shown to be the 56,980 Da polypeptide, and it appeared very likely that the 17,787 polypeptide was identical to the 19 kDa polypeptide. Not only was it of similar molecular weight but it appeared, from DNA sequence, to be cotranscribed with the 59 kDa polypeptide.

The structure of the clone appeared to consist of transcription promoter elements, a putative ribosome binding site, the coding region for the 59 kDa polypeptide, another ribosome binding site, the coding region for the 19 kDa polypeptide then an A-T rich region containing three putative transcription termination sites. This suggested that these two polypeptides were coded for by a bi-cistronic operon. Multicistronic operons are common in archaea, showing similarity to eubacteria in this respect (Brown *et al.*, 1989).

The sequences of these two polypeptides, the 59 kDa and 19 kDa, were used for database homology searches by the BLAST program (see Fig. 4.16 overleaf). The searches revealed homology with biotin carboxylases for the 59 kDa polypeptide and homology with biotin carboxyl carrier proteins for the 19 kDa polypeptide. Biotin carboxylase and biotin carboxyl carrier protein subunits are commonly constituents of acetyl-CoA carboxylase and other carboxylases. In addition, weaker homology was detected for the 59 kDa subunit with carbamoyl phosphate synthases and for the 19 kDa subunit with dihydrolipoamide acetyltransferases.

Figure 4.16 Summary of the results of protein homology searches. Extensive homology is 15-40 % identity, weak homology is 10-15 % identity.

Proteins showing extensive homology to 59 kDa

Escherichia coli biotin carboxylase (Li and Cronan, 1992; Kondo et al., 1991)

Pseudomonas aeruginosa biotin carboxylase (Best and Knauf, 1993)

Anabaena sp. biotin carboxylase (Gornicki *et al.*, 1993)

Proteins showing extensive homology to 19 kDa

Escherichia coli biotin carboxyl carrier protein (Muramatsu and Mizuno, 1989)

Pseudomonas aeruginosa biotin carboxyl carrier protein (Best and Knauf, 1993)

Anabaena sp. biotin carboxyl carrier protein (Gornicki *et al.*, 1993)

Propionibacterium freudenreichii trancarboxylase biotin carboxyl carrier protein (Wood and Kumov, 1985)

Salmonella typhimurium sodium ion pump oxaloacetate decarboxylase (α chain) (Woehlke *et al.*, 1992)

Klebsiella pneumoniae sodium ion pump oxaloacetate decarboxylase (α chain) (Schwarz *et al.*, 1988)

Tomato biotin binding fragment (Accession no. A29271)

Proteins showing extensive homology to both 59 and 19 kDa proteins

Mycobacterium leprae biotin carboxyl carrier protein (Accession no. S22188)

Mycobacterium tuberculosis biotin carboxyl carrier protein (Accession no. S31519)

Yeast urea carboxylase (Genbauffe and Cooper, 1991)

Human (Lamhonwah *et al.*, 1989), and rat (Browner *et al.*, 1989) propionyl-CoA carboxylase

Yeast (Lim *et al.*, 1988) and mouse (Zhang *et al.*, 1993) pyruvate carboxylase

Human, rat (López-Casillas *et al.*, 1988), chicken and yeast (Al-Feel *et al.*, 1992) acetyl-CoA carboxylase

Soybean 3-methylcrotonyl-CoA carboxylase (Song *et al.*, 1994)

Proteins showing weak homology to 59 kDa protein

Human, Rat, Hamster, Yeast, *Dictyostelium*, *Bacillus caldolyticus*, *Bacillus subtilis* and *Escherichia coli* carbamoyl phosphate synthetases (Nyunoya *et al.*, 1985).

Proteins showing weak homology to 19 kDa protein

Azotobacter vinelandii, *Escherichia coli*, *Bacillus stearothermophilus* and *Alcaligenes eutrophicus* dihydrolipoamide acetyltransferases (Ali and Guest, 1990).

4.18 - Biotin-dependent carboxylases

Biotin-dependent carboxylases can be subdivided into three classes (Knowles, 1989). Class I enzymes carboxylate biotin with bicarbonate in a reaction requiring ATP and Mg^{2+} then transfer the carboxyl group to acceptors such as acetyl-CoA, propionyl-CoA, pyruvate or urea. This group utilises the bicarbonate ion as substrate, in common with carbamoyl phosphate synthetases and phosphoenolpyruvate carboxylases. This is in contrast to enzymes such as RuBisCO which use dissolved CO_2 as a substrate.

Class II enzymes have been found only in anaerobic eubacteria where they mediate sodium transport, coupling this to the decarboxylation of β-keto acids and their thioesters. The only known example of a class III enzyme is the transcarboxylase from *Propionibacterium*, which couples two different carboxylation reactions.

Studies of the class I carboxylases have revealed that the two different reactions catalysed by these enzymes, the carboxylation of biotin and the transfer of the

carboxyl group from carboxybiotin to the acceptor molecule are carried out by different protein components each possessing the relevant active site. In addition another protein component is required to shuttle carboxybiotin between these two active sites.

The simplest example of this is the acetyl-CoA carboxylase of *E. coli*, where three separate proteins are responsible for these functions. These are a biotin carboxylase (BC), a biotin carboxyl carrier protein (BCCP), and a transcarboxylase (TC). In other organisms, however, these separate functions can be included on the same polypeptide, for example eukaryotic acetyl-CoA carboxylases possess all three activities on one polypeptide chain. Different domains may be identified as homologous to the separate proteins of *E. coli*, indicating a common evolutionary relationship for these proteins.

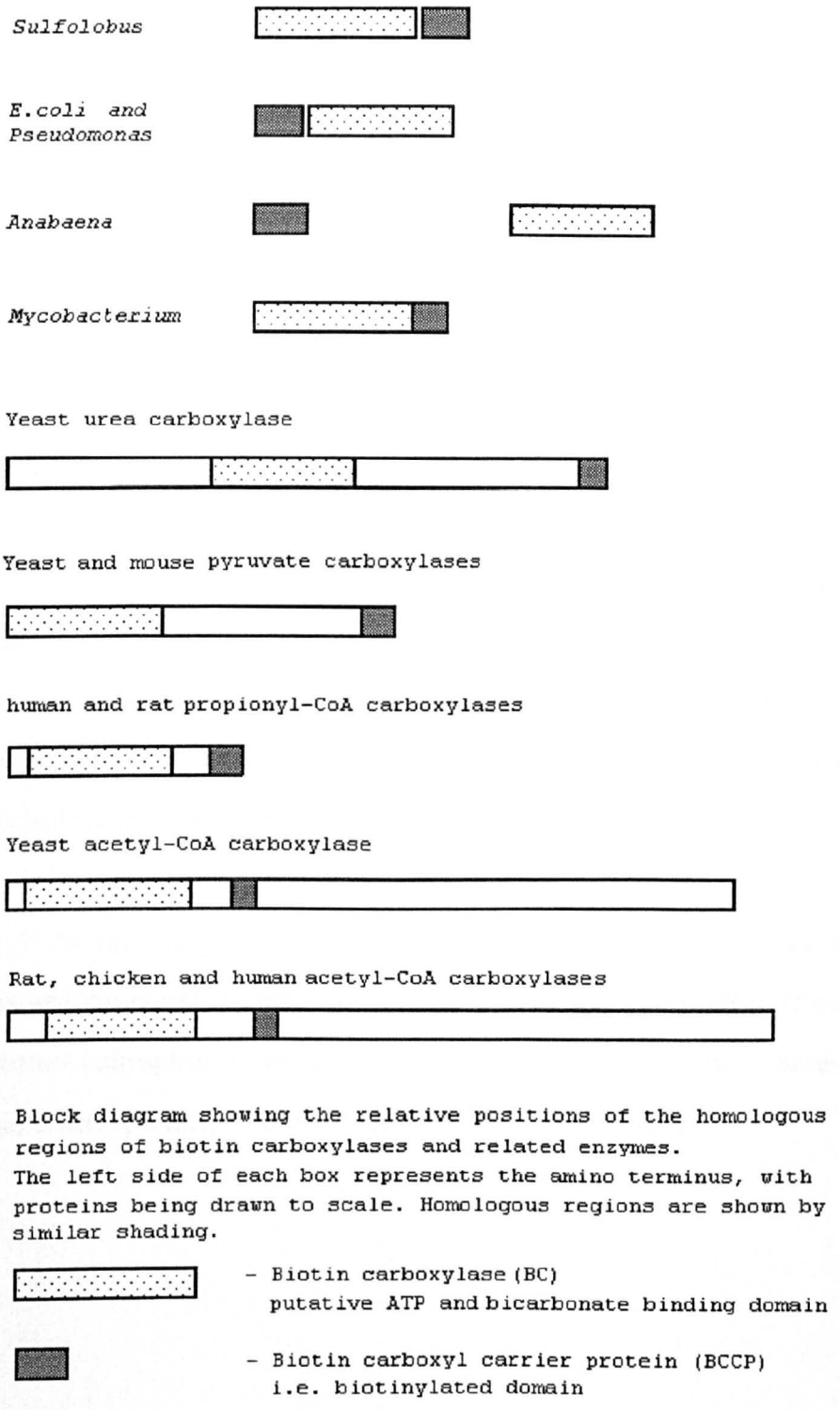
The organisation of the genes coding for acetyl-CoA carboxylases in eubacteria appears to vary. *E. coli* possesses four genes, *acc* A (the α subunit of transcarboxylase), *acc* D (the β subunit of transcarboxylase), *acc* B (biotin carboxyl carrier protein) and *acc* C (biotin carboxylase). Of these four genes, *acc* B and C are organised in one operon, *acc* A and D are found separately, unlinked to the *acc* BC operon (Li and Cronan, 1992) (Kondo *et al.*, 1991). A similar arrangement is found in *Pseudomonas* (Best and Knauf, 1993). *Anabaena* biotin carboxylase (BC) and biotin carboxyl carrier protein (BCCP) genes have also been isolated and found to be unlinked (Gornicki *et al.*, 1993).

Acetyl-CoA carboxylases from *Mycobacterium* appear to have a different structure, consisting of two subunits, one of which is biotinylated. The biotinylated subunit is termed the biotin carboxyl carrier protein, but includes both biotin carboxylase (BC) and biotin carboxyl carrier protein (BCCP) domains, possibly linked through a gene fusion event. Both the *Mycobacterium* acetyl-CoA carboxylase and

that of *Streptomyces* exhibit wider specificity than that of *E.coli*, in that they are both capable of carboxylating propionyl-CoA and *Streptomyces* acetyl-CoA carboxylase can carboxylate butyryl-CoA (Henrickson and Allen, 1979; Hunaiti and Kolattukudy, 1982).

Some of the different genetic organisations of the prokaryotes and protein domain structure of the eukaryotes are illustrated in figure 4.17 (overleaf).

Figure 4.17 Gene organisations of prokaryotic and protein domain structure of eukaryotic class I biotin dependent carboxylases.



4.19 - Comparisons between homologous proteins

Representatives of the amino acid sequences of the proteins found to be homologous to *Sulfolobus* LM biotin carboxylase (see Fig. 4.18) and biotin carboxyl carrier protein (see Fig. 4.20) were aligned.

The *Sulfolobus* putative biotin carboxylase sequence was aligned with sequences of the biotin carboxylase of *E. coli*, the biotin carboxyl carrier protein of *Mycobacterium* and biotin carboxylase domains of eukaryotic urea, pyruvate, acetyl-CoA and propionyl-CoA carboxylases as well as a carbamoyl phosphate synthetase (see Fig. 4.18).

The *Sulfolobus* putative biotin carboxyl carrier protein sequence was aligned with those of the biotin carboxyl carrier domains of *Mycobacterium* carboxylase, of *Propionibacterium* transcarboxylase, the oxaloacetate decarboxylase of *Salmonella typhimurium*, eukaryotic urea, pyruvate and propionyl-CoA carboxylases as well as the dihydrolipoamide acetyltransferase of *E. coli* (see Fig. 4.20).

In each case homology with the biotin related proteins may illustrate active sites and structural regions necessary for their functions whereas homology with the carbamoyl phosphate synthase and the dihydrolipoamide dehydrogenases may highlight regions less specifically related to biotin carboxylase activity.

4.20 - Identification of motifs within *Sulfolobus* biotin carboxylase

Comparisons between the sequence of *Sulfolobus* biotin carboxylase and those of different organisms show many homologous residues and conserved regions. This homology clearly shows that these proteins are related by common ancestry.

Various of these homologous regions have been identified as possible sites of substrate binding and activity. The biotin carboxylase subunit of *E. coli* is thought to possess sites for bicarbonate binding, ATP binding and an active site at which the biotinylated carrier protein (BCCP), or free biotin *in vitro*, is carboxylated (Li and Cronan, 1992).

The ATP binding site has been identified by comparison of biotin dependent carboxylases with carbamoyl phosphate synthetases (Lim *et al.*, 1988). Both these classes of enzymes bind ATP and bicarbonate. The glycine rich sequence centred at position 164 on the penultimate figure is similar to ATP binding sites of adenylate kinase and carbamoyl phosphate synthase, which have been confirmed by site directed mutagenesis studies (Reinstein *et al.*, 1988; Post *et al.*, 1989). This region in *Sulfolobus* is very highly conserved in comparison with biotin dependent carboxylases of different urkingdoms and carbamoyl phosphate synthetase. Therefore the assignment of an ATP binding site to this region seems straightforward.

The active site of the enzyme, however, is more difficult to assign with great confidence. Li and Cronan identified a cysteine residue (position 230) conserved throughout the class I biotin dependent carboxylases, but not present in carbamoyl phosphate synthetase or class II and III biotin dependent carboxylases which do not utilise bicarbonate. This cysteine is conserved in the *Sulfolobus* protein. The basis for accepting this as the active site is that a proton donor / acceptor group, typically a

thiol, is required for transfer of CO_2 from the postulated carboxyphosphate intermediate to biotin (Tipton and Cleland, 1988).

Kondo *et al.*(1991), however, postulated a conserved region between amino acids 290 and 294 of Glu-Xaa-Asn-Xaa-Arg as the active site of the *E.coli* enzyme. This region is conserved throughout the class I biotin dependent carboxylases and the carbamoyl phosphate synthetases and contains functional residues such as glutamic acid and arginine. This region is also conserved in the *Sulfolobus* protein.

The bicarbonate binding site of the enzyme is also unclear. Li and Cronan (1992) postulate that it is located adjacent to the glycine rich region around position 164, whereas Gornicki *et al.* (1993) consider it to be at cysteine 230.

The sequence of the *Sulfolobus* protein (see Fig. 4.18) would support any of the above models, but for any degree of certainty, site directed mutagenesis would have to be carried out. Alternatively, current X-ray crystallographic analysis of *E. coli* biotin carboxylase should elucidate the composition of the active site and various binding sites (Waldrop et al., 1994).

In summary, the amino acid sequence of *Sulfolobus* biotin carboxylase is in close agreement with consensus sequences throughout the group of class I biotin-dependent carboxylases providing strong evidence of its membership of this group of proteins.

Figure 4.18 Comparison of the amino acid sequences of biotin carboxylases and related proteins with the biotin carboxylase of *Sulfolobus* strain LM.

```
                1                                                  29
SBC     ..........  MPPFGKVLVS NRGEIAVSVM KRIKE.....  ......MGMK
EBC     ..........  ..MLDKIVIA NRGEIALRIL RACKE.....  ......LGIK
yPC     -AGLRDNFNL  LGEKNKILVA NRGEIPIRIF RTAHE.....  ......LSMQ
hPCC    -LGSVGYDPN  EKTFDKILVA NRGEIACRVI RTCKK.....  ......MGIK
MBCCP   .....MASHA  SSRIAKVLVA NRGEIAVRVI RAARH.....  ......ARLP
yUC     EMLKKKESQK  KKLFDTVLIA NRGEIAVRII KTLKK.....  ......LGIR
yACC    LRDFVKSHGG  HTVISKILIA NNGIAAVKEI RSVRKWAYET  FGDDRTVQFV
DCPS    ..........  .....KVLIL GSGGLSIGQA GEFDYSGSQA  IKALKEEGIK

        30                                                     76
SBC     AVAVYSEADK  YALHVKYADE AYYIGPA...  PSIESYLNIE  RIIDAAEKAH
EBC     TVAVHSSADR  DLKHVLLADE TVCIGPA...  PSVKSYLNIP  AIISAAEITG
yPC     TVAIYSHEDR  LSTHKQKADE AYVIGEVGQY  TPVGAYLAID  EIISIAQKHQ
hPCC    TVAIHSDVDA  SSVHVKMADE AVCVGPA...  PTSKSYLNMD  AIMEAIKKTR
MBCCP   SVAVYAEPDA  EAPHVRLADE AFALGGH...  TSAESYLDFG  KILDAAAKSG
yUC     SVAVYSDPDK  YSQHVTDADV SVPLHGT...  TAAQTYLDMN  KIIDAAKQTN
yACC    AMATPEDLEA  NAEYIRMADQ YIEVPGG...  TNNNNYANVD  LIVDIAERAD
DCPS    TILINPNIAT  VQTSPGLADK VYFLPVNAS.  .......SVQ  KVIENENPDG

        77                                                     124
SBC     ADAVHPGYGF  LSERADFAEA VEKA..GMTF  IGPSSEVMNR  VKSKLDGKEL
EBC     AVAIHPGYGF  LSENANFAEQ VERS..GFIF  IGPKAETIRL  MGDKVSAIAA
yPC     VDFIHPGYGF  LSENSEFADK VVKA..GITW  IGPPAEVIDS  VGDKVSARNL
hPCC    AQAVHPGYGF  LSENKEFARC LAAE..DVVF  IGPDTHAIQA  MGDKIESKLL
MBCCP   ANAIHPGYGF  LAENADFAQA VIDA..GLIW  IGPSPQSIRD  LGDKVTARHI
yUC     AQAIIPGYGF  LSENADFSDA CTSA..GITF  VGPSGDIIRG  LGLKHSARQI
yACC    VDAVWAGWGH  ASENPLLPEK LSQSKRKVIF  IGPPGNAMRS  LGDKISSTIV
DCPS    ILVTFGGQTA  LNCGIELYKS GILEKYNCKV  LGTPIETIIA  TEDRGIFAEK

         125                                                   149
SBC     .QRSRNAYSP  GSDGPVGSLD ..........  ..........  ...EALK.LT
EBC     MKKAGVPCVP  GSDGPLGDDM ..........  ..........  ...DKNRAIA
yPC     AAKANVPTVP  GTPGPIETVE ..........  ..........  ...EAL.DFV
hPCC    AKKAEVNTIP  GFDGVVKDAE ..........  ..........  ...EAVR.IA
MBCCP   AARAQAPLVP  GTPDPVKNAD ..........  ..........  ...EVV.AFA
yUC     AQKAGVPLVP  GS..LLITSV ..........  ..........  ...EEAKKVA
yACC    AQSAKVPCIP  WSGTGVDTVH VDEKTGLVSV  DDDIYQKGCC  TSPEDGLQKA
DCPS    LSEINERIAP  SMA.......  ..........  .........C  NSLEESLIEA

        150                                                    199
SBC     EKIGYPVMVK  AAYGGGGIGI TRADNPDQLM  DIWGRNQRLA  KEAFGRPDLY
EBC     KRIGYPVIIK  ASGGGGGRGM RVVRGDAELA  QSISMTRAEA  KAAFSNDMVY
yPC     NEYGYPVIIK  AAFGGGGRGM RVVREGDDVA  DAFQRATSEA  RTAFGNGTCF
hPCC    REIGYPVMIK  ASAGGGGKGM RIAWDDEETR  DGFRLSSQEA  ASSFGDDRLL
MBCCP   KEHGVPIAIK  AAFGGGGKGM KVARTLEEIS  ELYESAVREA  TVAFGRGECF
yUC     AELEYPVMVK  STAGGGGIGL QKVDSEEDIE  HIFETVKHQG  ETFFGDAGVF
yACC    KRIGFPVMIK  ASEGGGGKGI RQVEREEDFI  ALY....HQA  ANEIPGSPIF
DCPS    EKIGYPVIVR  AAYCLGGLGS GFADNKEQLT  ALVTEAMA..  ....TSSQVL
```

```
      200                                                        247
SBC   IEKAAVRPRH IETQLIGDKY GTYVVAFERE C..HNQRRNQ KLIEEAPSPA
EBC   MEKYLENPRH VEIQVLADGQ GNAIYLAERD C..SMQRRHQ KVVEEAPAPG
yPC   VERFLDKPKH IEVQLLADNH GNVVHLFERD C..SVQRRHQ KVVEVAPAKT
hPCC  IEKFIDNPRH IEIQVLGDKH GNALWLNERE C..SIQRRNQ KVVEEAPSIF
MBCCP VERYLDKPRH VEAQVIADQH GNIVVAGTRD C..SLQRRFQ KLVEEAPAPF
yUC   LKRFIENARH VEVQLMGDGF GKAIALGERD C..SLQRRNQ KVIEETPAPN
yACC  IMKLAGRARH LEVQLLADQY GTNISLFGRD C..SVQRRHQ KIIEEAPVTI
DCPS  VEKSLKGWKE IEYEVLRDSK DNCITVCNME NFDPLGIHTG ESIVVAPSQN

      248                                                        297
SBC   LTWEKREEFI DASVKYGKKI GYFALGTMEF AYSDTTKDLY FLELNKRLQV
EBC   ITPELRRYIG ERSRKACVDI GYRGAGTFEF LF.E.NGEFY FIEMNTRIQV
yPC   LPREVRDAIL TDAVKLAKEC GYRNAGTAEF LVDNQNRH.Y FIEINPRIQV
hPCC  LDAETRRAMG EQAVALARAV KYSSAGTVEF LV.DSKKNFY FLEMNTRLQV
MBCCP LTDAQRKEIH ESAKRICKEA HYYGAGTVEY LV.GQDGLIS FLEVNTRLQV
yUC   LPEKTRLALR KAAESLGSLL NYKCAGTVEF IYDEKKDEFY FLEVNTRLQV
yACC  AKAETFHEME KAAVRLGKLV GYVSAGTVEY LYSHDDGKFY FLELNPRLQV
DCPS  LSDREYQMLR ETAIKTVRHL GVIGECNIQY SLNPYSEEYC IIEVNARLSR

      298                                  326
SBC   EHGITELV.. ....TGIDLV KLQIRLVAGE HMP.F..... .........
EBC   EHPVTEMI.. ....TGVDLI KEQLRIAAGQ PLS.I..... .........
yPC   EHTITEEI.. ....TGIDIV AAQIQIAAGA SLPQL..... .........
hPCC  EHPVTECIHW PGPSPGKTVL QEHLSGTNKL IFA.F..... .........
MBCCP EHPVTE.... ..ETTGIDLV LQQFKIANGE KLE.L..... .........
yUC   EHPITEMV.. ....TGLDLV EWMIRIAAND APDF..... .........
yACC  EHPTTEMV.. ....SGVNLP AAQLQIAMGI PMHRISDIRT LYGMNPHSAS
DCPS  SSALAS.... ..KATGYPLA FISAKVALGY DLAAL..... .........

                  327                                        364
SBC   .......... .TQEELKKRL RGHAIEYRIN AEDPLNNFT. GSSGFVTYYK
EBC   .......... .KQEEV..HV RGHAVECRIN AEDPNTF..L PSPGKITRFH
yPC   .......... .GLFQDKITT RGFAIQCRIT TEDPAKNFQ. PDTGRIEVYR
hPCC  .......... .N........ .GWAVECRVY AEDPYKSFGL PSIGRLSQYQ
MBCCP .......... .IKDPI..PC .GHAIEFRIN GEDAGRNF.L PSPGPVSKFH
yUC   .......... ...DSTKVEV NGVSMEARLY AENPLKNFR. PSPGLLVDVK
yACC  EIDFEFKTQD ATKKQRRPIP KGHCTACRIT SEDPNDGFK. PSGGTLHELN
DCPS  .......... .......... RNTITKKTTA CFEPSLDYLV VKMPRWDLKK

      365                                                        412
SBC   EPSG.PGVRL DTG.IMEGRY VPPFYDSLVA NLMVYGENRS GAIEVGRRSL
EBC   APGGF.GVRW ESH.IYAGYT VPPYYDSMIG KLICYGENRD VAIARMKNAL
yPC   SAGG.NGVRL DGGNAYAGTI ISPHYDSMLV KCSCSGSTYE IVRRKMIRAL
hPCC  EPLHLPGVRV DSG.IQPGSD ISIYYDPMIS KLITYGSDRT EALKRMADAL
MBCCP PPTG.PGVRL DSG.VETGSV IGGQFDSMLA KLIVHGATRQ EALARARRAL
yUC   FPDW...ARV DT.WVKKGTN ISPEYDPTLA KIIVHGKDRD DAISKLNQAL
yACC  FRSSSNV..W GYFSVGNNGN IHSFSDSQFG HIFAFGENRQ ASRKHMVVAL
DCPS  FTRVSNKISS SMKSVGEVMS IGRKFEEAIQ KAI....... ..RMVMDGAV

      413                                                        458
SBC   RDLQIGG.IK TTIPLYKLIM DDEDYQNGNF TTAYIAD... KSEIFTKKLR
EBC   QELIIDG.IK TNVDLQIRIM NDENFQHGGT NIHYL..... .....EKKLG
yPC   IEFRIRG.VK TNIPFLLTLL TNPVFIEGTY WGTFIDD.TP QLFQMVSSQN
hPCC  DNYVIRG.VT HNIALLREVI INSRFVKGDI STKFLSD.VY PDGFKGHMLT
MBCCP DEFEVEG.LA TVIPFHRAVV SDPALIGDNN SFSVHT.... ..........
yUC   EETKVYG.CI TNIDYLKSII TSDFFAKAKV STNILNSYQY EPTAIEITLP
yACC  KELSIRGDFR TTVEYLIKLL ETEDFEDNTI TTGWLDDLIT HKMTAEKPDP
DCPS  EGFQAGVFPT SDEELEHPTN NRILVLASAF KDGYSIDRVH QLTKIDKWFL
```

```
      459                                                        502
SBC   E....KEMLK TALAASVYNR GLLRGSSGNN NVQAQDNGKP RSAW..KTYG
EBC   L....QEK.. .......... .......... .......... ..........
yPC   R....AQKLL HYLADVADNG SSIKGQIGLP KLKSNPSVPH LHDA..QGNV
hPCC  K....SEKNQ LLAIASSLFV AFQLRAQHFQ ENSRMPVIKP DIAN..WELS
MBCCP .......... .......... .......RWI ETEWNNTIEP FIDN..QPLD
yUC   G....AHTSI QDYPGRVGYW RIGVPPSGPM DAYSFRLANR IVGNDYRTPA
yACC  TLAVICGAAT KAFLASEEAR HKYIESLQKG QVLSKDLLQT MFPVDFIHEG
DCPS  TKLKAIIDLE NHLSTYKEPS QIPSEILKFS KQQGFSDKQI ARAVGTTELN

      503       513
SBC   VQAQAPARV. MW........
EBC   .......... ..........
yPC   INVTKSAPPS GWRQVLLEK-
hPCC  VKLHDKVHT. VVASNNGSV-
MBCCP EEDTRPQQT. VI........
yUC   IEVTLTGPSI VFHCETVIA-
yACC  KRYKFTVAKS GNDRYTLFI-
DCPS  VRDYRKKMGI IPCTKHIDTV
```

SBC - biotin carboxylase of *Sulfolobus* LM
EBC - biotin carboxylase of *Escherichia coli*
yPC - yeast pyruvate carboxylase 1(from amino acids 7-543)
hPCC - human propionyl CoA carboxylase alpha chain precursor (from amino acids 26-583)
MBCCP - biotin carboxyl carrier protein of *Mycobacterium leprae* (upto amino acid 536)
yUC - yeast urea carboxylase (from amino acids 672-1152)
yACC - yeast acetyl CoA carboxylase (from amino acids 47-647)
DCPS - carbamoyl-phosphate synthase of slime mold *Dictyostelium discoideum* (upto amino acid 515)

Residues identical to those of the *Sulfolobus* biotin carboxylase are shown in bold. Numbering is for the *Sulfolobus* protein.

Figure 4.19 Percentage identity of *Sulfolobus* biotin carboxylase amino acids with the relevant domains of related proteins.

E. coli biotin carboxylase	- 39 %
yeast pyruvate carboxylase	- 36 %
human propionyl-CoA carboxylase	- 35 %
Mycobacterium biotin carrier protein	- 35 %
yeast urea carboxylase	- 33 %
yeast acetyl-CoA carboxylase	- 29 %
carbamoyl phosphate synthetase	- 13 %

4.21 - Identification of motifs within the biotin carboxyl carrier protein

The putative biotin carboxyl carrier protein of *Sulfolobus* does not exhibit such extensive homology with related proteins (see Fig. 4.20) as does the *Sulfolobus* biotin carboxylase. However, it does show a high degree of homology with biotin carboxyl carrier proteins over a short region between amino acids 110 and 150.

The N-terminal portions of the proteins compared above, including the *Sulfolobus* protein, show no great similarity to one another. Proteolytic cleavage of this area produces truncated peptides which still bind biotin, but interact much more weakly with biotin carboxylase and transcarboxylase subunits than full length BCCP (Fall *et al.*, 1975). This suggests that the N-terminal part of the protein is the site of association with the companion biotin carboxylase and transcarboxylase proteins. The lack of homology between different biotin carboxyl carrier proteins in this area may be explained by this but requires the BCCP binding sites of the biotin carboxylases and transcarboxylases to differ. As the locations of these sites have not yet been identified this would appear possible.

The region between amino acids 110 and 150 contains, from amino acid 130, a sequence absolutely conserved between the biotin carboxyl carrier proteins. This is -Ala-Met-Lys- (AMK). The similar sequence motif -Ala-Met-Lys-Met- (AMKM) has been identified in many biotin dependent enzymes as a biotin binding site (Morris *et al.*, 1987), the biotin being bound to the lysine residue (Wood and Barden, 1977). It appears that the second methionine may not be essential for biotin binding as neither *Sulfolobus* biotin carboxyl carrier protein nor yeast urea carboxylase possess this residue in this position, non-conservative substitutions to Ala and Ser having perhaps occurred.

In the 110-150 amino acid region of *Sulfolobus* biotin carboxyl carrier protein there are a number of similarities to dihydrolipoamide dehydrogenases as well as biotin carboxylases. This has been interpreted as evidence both of evolutionary relatedness and similarity of structure relating to function (Knowles, 1989). The similar function of both classes of proteins is that they are both thought to rely on a 'swinging arm' moiety for transfer of cofactors (lipoic acid or biotin) between specific sites of an enzyme complex. The lipoic acid attachment point is thought to be in a similar position to that of biotin. Therefore it appears that the surrounding homologous amino acids may contribute to the flexibility of this part of the molecule.

This locality is particularly rich in proline, alanine, glycine and valine residues. The alanine and proline residues have been proposed as the source of flexibility as alanine / proline rich regions are known to be highly mobile (Radford *et al.*, 1989). However, the sequences above differ markedly in the quantity and position of alanine and proline residues. There are a number of glycine, valine and alanine residues conserved between the majority of the aligned proteins. It would appear likely that these short chain aliphatic residues are required for the structural integrity of the area surrounding the site of biotinylation. The presence of a large region of proline and alanine residues would perhaps confer a degree of mobility that is not required for the function of a majority of the group of biotin carboxylases as the distance between the active sites of transcarboxylase has been estimated as only 7 Å (Knowles, 1989) whereas the length of a biotinylated lysine residue is known to be 14 Å . However, the extent to which high mobility of the biotinylated lysine is necessary may vary between carboxylases if the distance between biotin carboxylase and carboxyl transferase active sites varies. Indeed, high mobility of protein segments may be disadvantageous in terms of protein stability for thermophilic enzymes such as those from *Sulfolobus*.

Figure 4.20 Comparison of the amino acid sequences of biotin carboxyl carrier and related proteins with putative biotin carboxyl carrier protein of *Sulfolobus* strain LM.

```
SBCCP  ....MKLLRV YMETGETYIA SYDQKDNKDT VKMEEGELNV EFLGRGTREN  46
yUC    -LCLAASSEV PWLMNPFDQV EFYPVSEEDL DKMTEDCDNG VYKVNIEKSV
PBCCP  .......... .......... .......... ....MKLK.. ..VTVNGTAY
yPC    -YNMYPRVYE DFQKMRETYG DLSVLPTRSF LSPLETDEEI EVVIEQGKTL
StOD   -VALFPQIGL KFLENRNNPA AFEPLPQAEA AQPVAKAE.. ..KPAASGIY
MBCCP  -QPLDEEDTR PQQTVI.... .....VEVDG RRLEVSLPAD LA........
hPCC   -WELSVKLHD KVHTVVASNN GSVFSVEVDG SKLNVTSTWN LASPLLSVSV
EDA    .......... .......... .......... .......... ..........

SBCCP  EYLFKVGNEV HSITIDRGFL ILDQEEEYKV DRITELPVKE GQSVEELM..  94
yUC    FDHQEYLRWI NANKDSITAF QEGQLGERAE EFAKLIQNAN SELKESVTVK
PBCCP  DVDVDVDKSH ENPMG.TILF GGGTGGAP.. ..APRAAGGA GAGKAGEG..
yPC    IIKLQAVGDL NKKTGEREVY FDLNGEMRKI RVADRSQKVE TVTKSKAD..
StOD   TVEVEGKAFV VKVSD.GGDI SQLTAAVP.. ..AASSAPVQ AAAPAGAG..
MBCCP  .........L ANPAGCN... .........PA GVIRKKPKPR KRGGHTGA..
hPCC   DGTQRTVQCL SREAGGNMSI QFLGTVYKVN ILTRLAAELN KFMLEKVT..
EDA    .......... .......... .......... .......... ..........

SBCCP  .......KGK EGEVLSPLQG RVVAIRVKEG DAVTKGQPLL SVEAMKSETI 137
yUC    PDEEEDFPEG AEIVYSEYSG RFWKSIASVG DVIEAGQGLL IIEAMKAEMI
PBCCP  .......E.. ...IPAPLAG TVSKILVKEG DTVKAGQTVL VLEAMKMETE
yPC    .......MHD PLHIGAPMAG VIVEVKVHKG SLIKKGQPVA VLSAMKMEMI
StOD   .......T.. ..PVTAPLAG NIWKVIATEG QTVAEGDVLL ILEAMKMETE
MBCCP  .......ATS GDAVTAPMQG TVVKVAVAEG QTVMTGDLVV VLEAMKMENP
hPCC   .......EDT SSVLRSPMPG VVVAVSVKPG DAVAEGQEIC VIEAMKMQNS
EDA    .....AIEIK VPDIGADEVE .ITEILVKVG DKVEAEQSLI TVEGDKASME

SBCCP  ISAPIAGVIE KIAVKTRPGS KEEETCWLY. .......... .......... 166
yUC    ISAPKSGKII KICHGNGDMV DSGDIVAVIE TLADD..... ..........
PBCCP  INAPTDGKVE KVLVKERDAV QGGQGLIKIG .......... ..........
yPC    ISSPSDGQVK EVFVSDGENV DSSDLLVLLE DQVPVETKA. ..........
StOD   IRAAQAGTVR GIAVKSGDAV SVGDTLMTLA .......... ..........
MBCCP  VTAHKDGIIT GLAVEAGTAI TQGTVLAEIK .......... ..........
hPCC   MTAGKTGTVK SVHCQAGDTV GEGDLLVELE .......... ..........
EDA    VPAPFAGVVK ELKVNVGDKV KTGSLIMIFE VEGAAPAAAP A.........
```

SBCCP - putative biotin carboxyl carrier protein of *Sulfolobus* LM
yUC - yeast urea carboxylase (from amino acid 1653)
PBCCP - biotin carboxyl carrier protein of *Propionibacterium freudenreichii* subsp. shermanii
yPC - yeast pyruvate carboxylase 1 (from amino acid 990)
StOD - sodium ion pump oxaloacetate decarboxylase subunit alpha of *Salmonella typhimurium* (from amino acid 455)
MBCCP - biotin carboxyl carrier protein of *Mycobacterium leprae* (from amino acid 465)
hPCC - human propionyl CoA carboxylase alpha chain precursor (from amino acid 530)
EDA - lipoyl domain of the dihydrolipoamide acetyltransferase of *Escherichia coli* pyruvate dehydrogenase complex E2p

Residues identical to those of the *Sulfolobus* protein are shown in bold. Numbering is for the *Sulfolobus* protein.

Figure 4.21 Percentage identity of *Sulfolobus* putative biotin carboxyl carrier protein with relevant domains of related proteins.

yeast urea carboxylase	21 %
Propionibacterium BCCP	19 %
yeast pyruvate carboxylase	17 %
Salmonella oxaloacetate decarboxylase	17 %
Mycobacterium biotin carrier protein	17 %
human propionyl-CoA carboxylase	17 %
E. coli dihydrolipoamide acetyltransferase	13 %

4.22 - Evolutionary relationships between biotin carboxylases

Amino acid sequences of seventeen biotin carboxylases or domains of carboxylases showing homology to biotin carboxylases were compared using GCG computer programs. Sequences were aligned, alignment was optimised, pairwise distance matrices were calculated using three different methods and two variants of dendrograms of relatedness were drawn for each matrix. The same procedure was followed for nineteen biotin carboxyl carrier proteins and biotinylated domains of carboxylases.

For the biotin carboxylases a 350 residue region of high homology corresponding to amino acids 6 to 356 of figure 4.18 was aligned. For the biotin carboxyl carrier proteins a shorter sequence of 50 residues from positions 100 to 150 of figure 4.20 was aligned as the N-terminal portions of these proteins show little homology to one another. For phylogenetic analysis, it is assumed that residues in the same position in two sequences are derived from an ancestral residue at the same position. Therefore, only regions of high similarity may be compared. Difference matrices derived for these regions do not give values pertaining to the whole protein, but only to these short regions. Alignments were checked and corrected manually.

The phylogenetic trees produced by these methods were compared with one another and comparisons made between the trees for biotin carboxylases and those for biotin carboxyl carrier proteins. Major differences were seen in the deep branching patterns of trees produced using different methods, thus showing that little confidence may be placed in evolutionary prediction from the data in this case. There are a number of factors which may be responsible for this.

The alignment of sequences is crucial for comparisons of relatedness between proteins. As the alignment produced by the software may not be optimal, manual alignment of the data output is recommended. Manual alignment is necessarily subjective and is unlikely to result in a truly optimal alignment. The nature of the proteins being compared also has a bearing on the results. Although an accurate description of the similarity between proteins may be possible, minor differences in structure and function between the proteins may alter the position of proteins on an evolutionary tree. In this case the function of each protein is assumed to be similar to those of others of the group, whereas it would appear likely that the different nature of interactions between subunits or domains of the carboxylases may affect parts of the compared sequences. In addition physical effects such as temperature optima and pH at the site of action would be expected to skew results. The effect of these differences is that sequence divergence becomes too high for determination of reliable phylogenetic trees.

Although the order of deep branches from the trees could not be relied upon, the clusters of proteins linked by lesser branches were reproducible when matrices were calculated using a range of different methods.

For the biotin carboxylases five different groups were identified (see Fig. 4.22). For the comparison between biotin carboxyl carrier proteins similar groups were seen (see Fig 4.22). However, the addition of eubacterial type II oxaloacetate decarboxylase

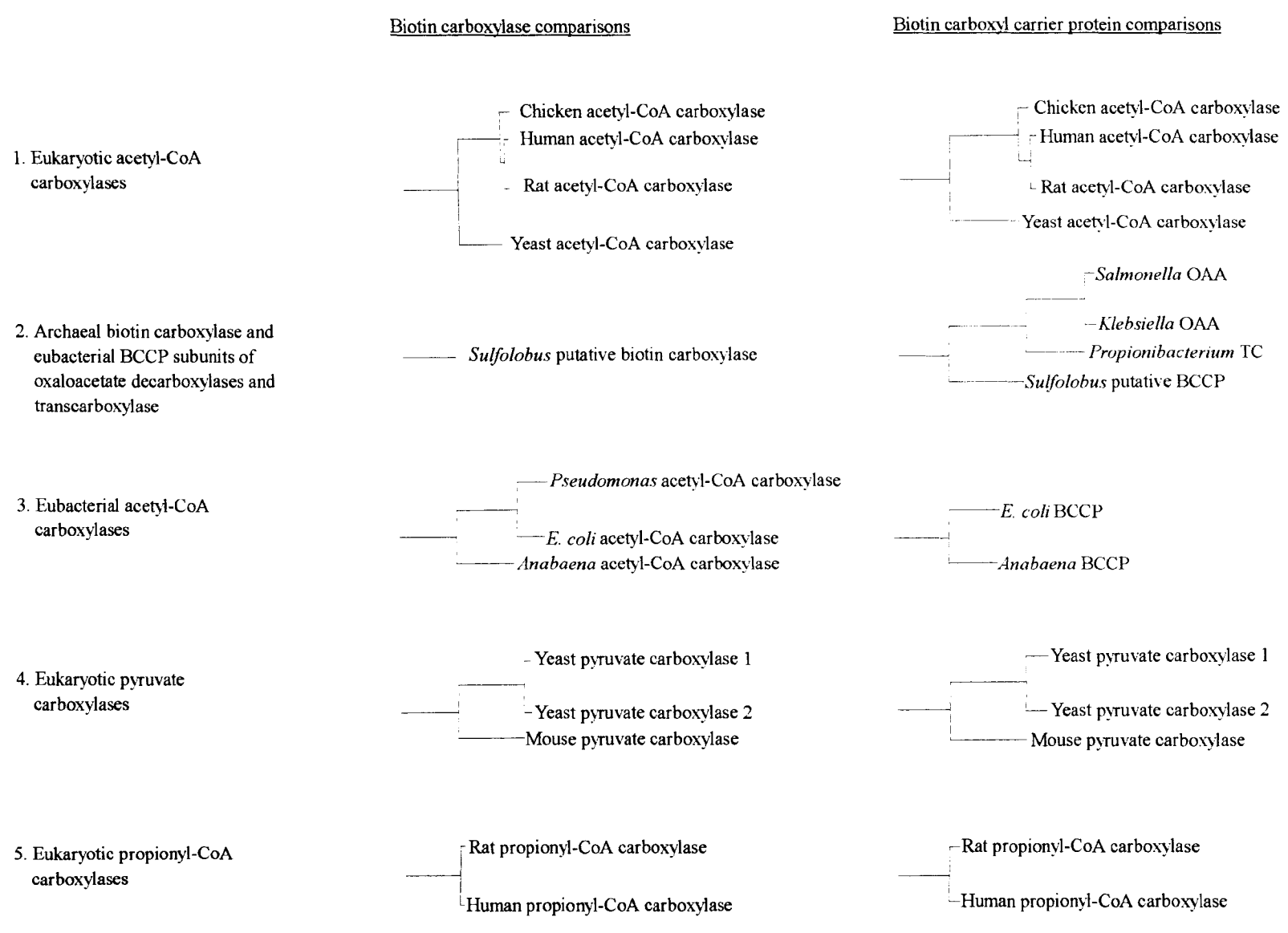
BCCPs and of eubacterial type III transcarboxylase BCCP was possible in this comparison, none of these proteins possessing a domain homologous to biotin carboxylase. Interestingly, these proteins clustered with the *Sulfolobus* protein.

Phylogenetic comparison of the carboxyl transferase subunits of various carboxylases has yielded a preliminary evolutionary classification of biotin carboxylases (Toh *et al.*, 1993). Toh's study detected the early divergence of the eukaryotic type acetyl-CoA carboxylases then divergence of eukaryotic type propionyl-CoA carboxylases, the remaining group consisted of pyruvate carboxylases and eubacterial oxaloacetate decarboxylases, acetyl-CoA carboxylases and transcarboxylase.

The results of my phylogenetic studies broadly agreed with this scheme. The early divergence of the five groups of carboxylases is apparent, although the points of divergence of the various groups of carboxylases is unclear. Toh *et al.*, (1993) concluded that the ancestral type of biotin-dependent carboxylase possessed carboxyl transferase which used acyl-CoA as the CO_2 acceptor and possessed biotin carboxylase with biotin carboxyl carrier protein. The divergence of the eukaryotic type acetyl-CoA and propionyl-CoA carboxylases was postulated to have occurred before the divergence of eubacteria and eukaryotes. My data would tend to support this hypothesis and extend it to before the divergence of archaea from the other urkingdoms. Indeed the possible inclusion of the *Sulfolobus* biotin carboxyl carrier protein with the oxaloacetate decarboxylases and transcarboxylase may indicate that the majority of functional changes between these proteins occurred early in evolution.

Figure 4.22 (Overleaf) Dendrograms of five major groups of carboxylases identified by phylogenetic analysis of biotin carboxylase proteins and of biotin carboxyl carrier proteins or homologous domains. These were produced from a Kimura distance matrix using the UPGMA method.

Figure 4.22 Dendrograms of five major groups of biotin carboxylase and biotin carboxyl carrier proteins



Biotin dependent acetyl-CoA carboxylases are found in eubacteria and eukaryotes catalysing the first committed step of fatty acid synthesis, the carboxylation of acetyl-CoA to form malonyl-CoA (Lane *et al.*, 1974). The biotin carboxylase from autotrophically grown *Sulfolobus* may not necessarily perform the same function.

The biotin carboxylase from *Sulfolobus* LM has been implicated in CO_2 fixation in the presence of ATP and acetyl-CoA. It is likely that acetyl-CoA is the physiological substrate of this enzyme. However, this study did not result in the isolation or characterisation of the carboxyl transferase, the portion which determines the specificity of the overall carboxylation reaction.

Apart from acetyl-CoA carboxylation the only other CO_2 fixation activity found in cell free extracts was that of PEP carboxylase (Norris *et al.*, 1989). What was striking about this study was that, in terms of the magnitude of CO_2 fixation, the ATP-dependent acetyl-CoA carboxylase activity was similar to the PEP carboxylase activity, though present in different sucrose gradient fractions, and no other fixation of CO_2 of comparable magnitude was detected. Fixation of CO_2 during autotrophic growth in *Acidianus brierleyi* and *Thermoproteus neutrophilus* has been shown to proceed via the action of a reductive carboxylic acid cycle (Kandler and Stetter, 1981; Schafer *et al.*,1986). It has been proposed that acetyl-CoA is carboxylated to pyruvate which proceeds to PEP which is itself carboxylated, re-entering the reductive citric acid cycle at oxaloacetate.

Initially an attractive hypothesis may be made for the biotin carboxylase of *Sulfolobus* being part of an acetyl-CoA carboxylase responsible for the production of pyruvate. The pyruvate would then be carboxylated to oxaloacetate by the PEP carboxylase noted above. This would explain the dramatic induction of biotin carboxylase seen under CO_2 limiting conditions, as acetyl-CoA carboxylases involved in fatty acid synthesis are not seen to be expressed to these levels. Indeed, the acetyl-

CoA carboxylase would appear to be the key enzyme of CO_2 assimilation during autotrophic growth in *Sulfolobus*.

However there are a number of unresolved observations which may cast doubt on this hypothesis. These are based on ^{14}C fixation experiments carried out using whole cell extracts of *Sulfolobus* LM (Norris et al., 1989). If pyruvate was indeed the product of the carboxylation of acetyl-CoA, labelled alanine would be expected as one of the first products. This was not found to be the case. Two as yet unidentified products were produced first, followed by malate, aspartate and glutamate before labelled PEP and phosphoglyceric acid. Clearly an extended biochemical study of CO_2 fixation may in future provide the resolution of its pathway and clarify the role of the biotin carboxylase during conditions of limited CO_2.

CHAPTER 5

A protein induced during iron oxidation in *Sulfolobus* strain LM

5.1 - Introduction

Various species of thermoacidophilic archaea oxidize sulphur compounds and ferrous iron and are capable of autotrophic growth on these substrates. In contrast to well studied iron-oxidising eubacteria such as *Thiobacillus ferrooxidans* and *Leptospirillum ferrooxidans*, nothing is known of the mechanism of iron oxidation in archaea. In addition, little is known of the mechanisms by which the thermoacidophilic archaea survive high concentrations of metal ions.

Zinc- and copper-induced proteins have been noted in preliminary work with the non-iron-oxidising *Sulfolobus solfataricus* and the eubacterial thermoacidophile, *Bacillus (Alicyclobacillus) acidocaldarius* (Scudiero et al., 1992), but details of the proteins were not given.

In order to isolate and characterise proteins related to growth on ferrous iron, a comparative approach was taken whereby proteins induced during iron oxidation were identified and further characterised. A molecular biological approach was used to initiate the study of these proteins in order to avoid the problems of having to isolate and purify the proteins from the very limited biomass produced during growth on ferrous iron. For this work *Sulfolobus* LM was the species chosen because its autotrophic growth has been well described (Norris *et al.*, 1986; Wood *et al.*, 1987; Nixon and Norris, 1989) and it is the subject of much work on mineral sulphide oxidation (Norris and Owen, 1992; LeRoux and Wakerley, 1988; Liu *et al.*, 1991), for which iron oxidation is the critical step.

5.2 - Autotrophic growth of *Sulfolobus* LM

In order to identify proteins induced during iron oxidation, it was necessary to compare cells grown by iron oxidation with those grown without the oxidation of iron. The simplest method of carrying this out was to grow cells autotrophically using ferrous iron and autotrophically using tetrathionate.

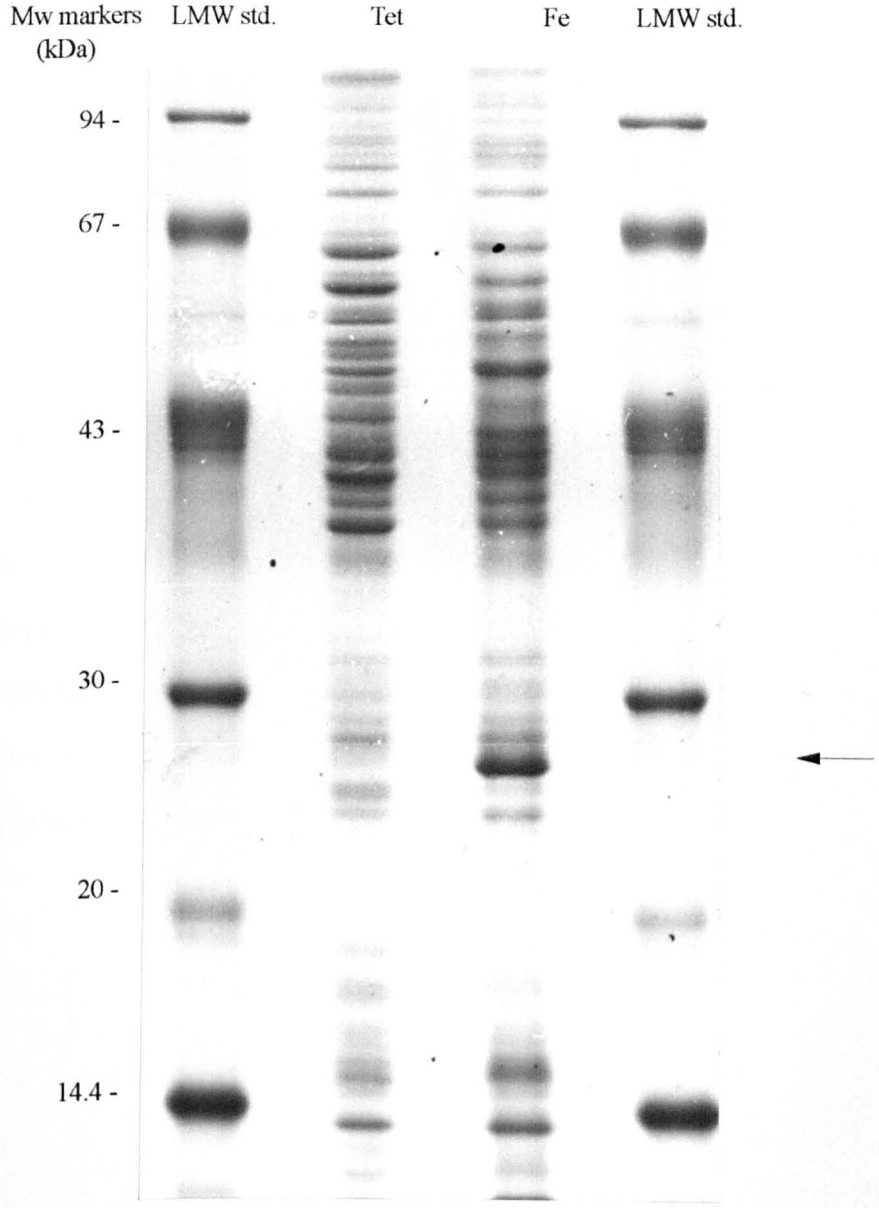
Sulfolobus LM was obtained from stocks grown using sulphur. Ten serial subcultures were carried out with cells grown on tetrathionate medium, and ten with cells grown on ferrous iron medium. This large number of serial subcultures was necessary in order to remove all traces of elemental sulphur from the medium, and to induce possible iron oxidation related proteins.

5.3 - SDS-PAGE of whole cell proteins of iron and tetrathionate-grown cells

SDS-PAGE was carried out using samples of iron-grown and tetrathionate-grown *Sulfolobus* LM. A 10% (w/v) acrylamide gel was used and protein bands were visualised by Coomassie staining. Gels were examined for the presence of increased proportions of protein bands in the samples grown using ferrous iron.

This comparison resulted in the identification of a band of approximately 27 kDa apparent molecular mass on SDS-PAGE which was clearly present in greater concentration in ferrous iron-grown cells than in tetrathionate-grown cells (see Fig. 5.1). This observation was reproducible over a number of SDS-PAGE comparisons.

Figure 5.1 10% SDS-PAGE of lysates of *Sulfolobus* LM cells grown using ferrous iron (Fe) and using tetrathionate (Tet), with low molecular weight standards (LMW std.). A band at 27 kDa in the iron-grown cells is indicated.



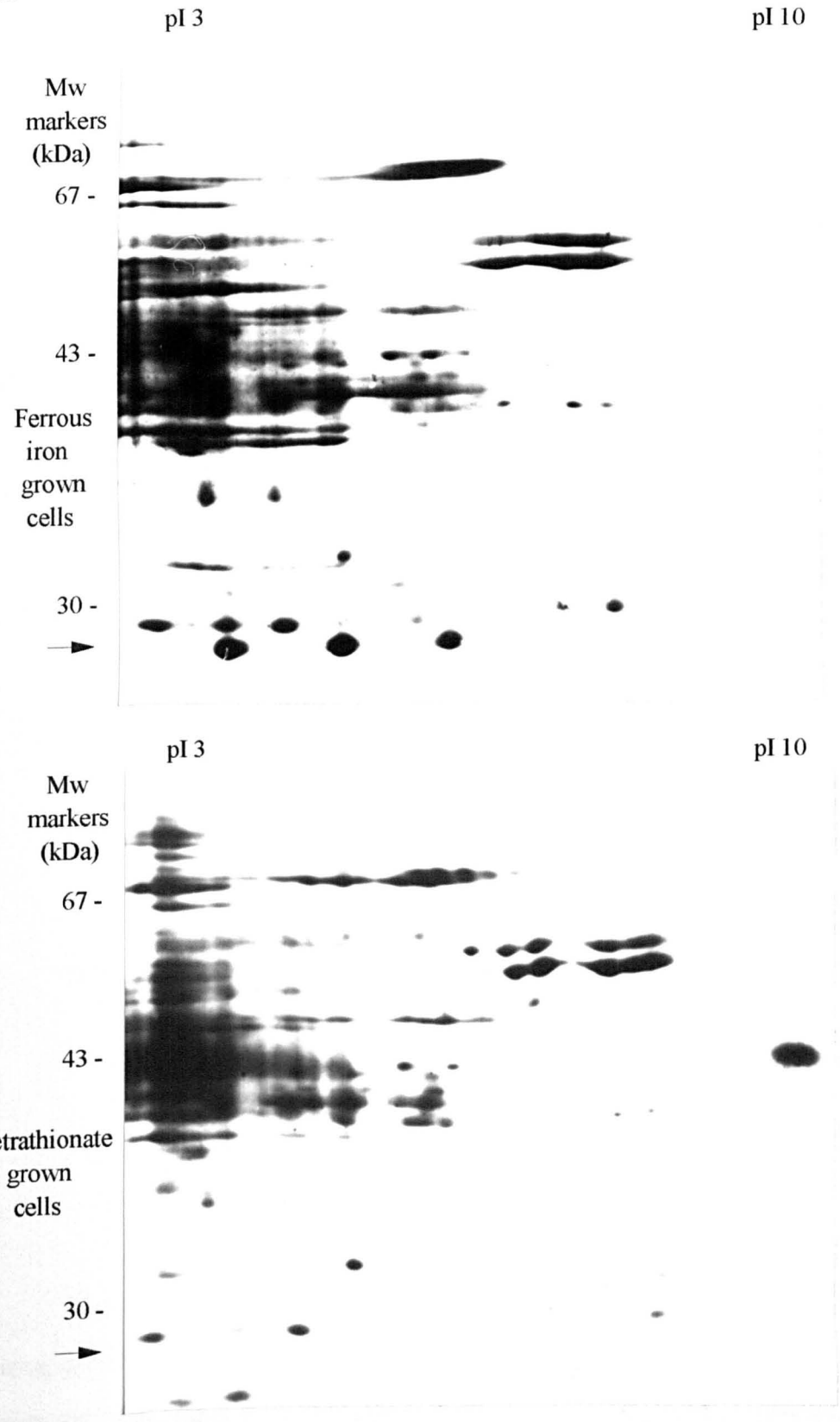
5.4 - 2D-PAGE investigation of the 27 kDa polypeptide

2D-PAGE was carried out using lysates of tetrathionate-grown cells and of iron-grown cells using IEF tube gels with a pI range of 3-10 for the first dimension and an SDS polyacrylamide slab gel for the second dimension (see Fig. 5.2 overleaf).

Electrophoresis of iron-grown cells showed the presence of three spots of equal intensity running at 27 kDa, but showing different isoelectric points. On the gel of tetrathionate-grown cells these spots were visible at similar positions, but were of far lower intensity. These spots presumably represent the 27 kDa polypeptide noted previously by SDS-PAGE.

The reason for three iron-induced polypeptide bands of different isoelectric points was not clear. They could have represented three entirely unrelated iron-induced proteins, but it appeared unlikely that these should all share the same molecular weight as estimated by SDS-PAGE. More likely, perhaps, was the presence of three differing forms of isoenzymes possessing similar but not identical characteristics. These would have been expected to be produced by different genes. Alternatively, different post translational modifications may have been responsible.

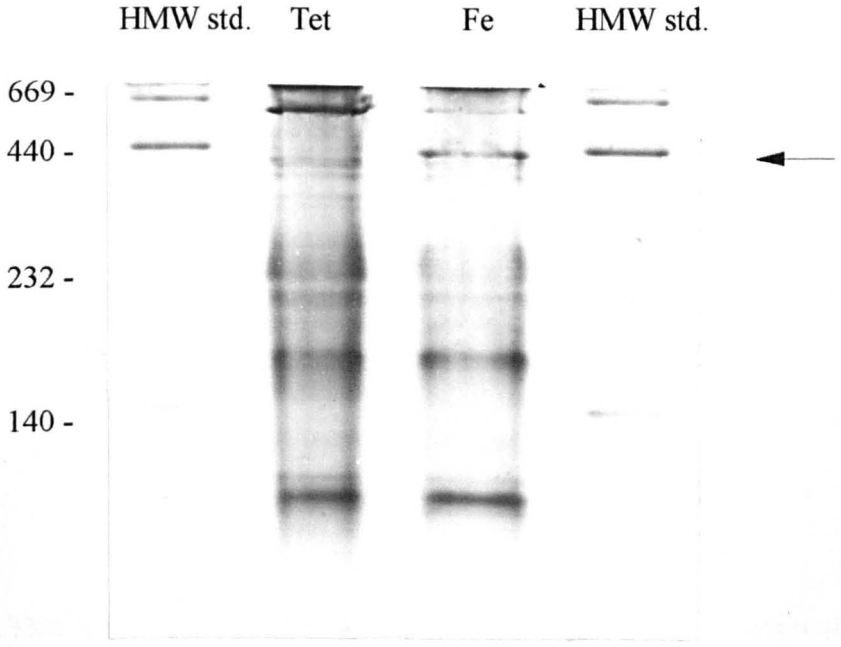
Figure 5.2 2D-PAGE of lysates of iron-grown and tetrathionate-grown *Sulfolobus* LM



5.5 - Native PAGE comparisons of tetrathionate- and iron-grown cells

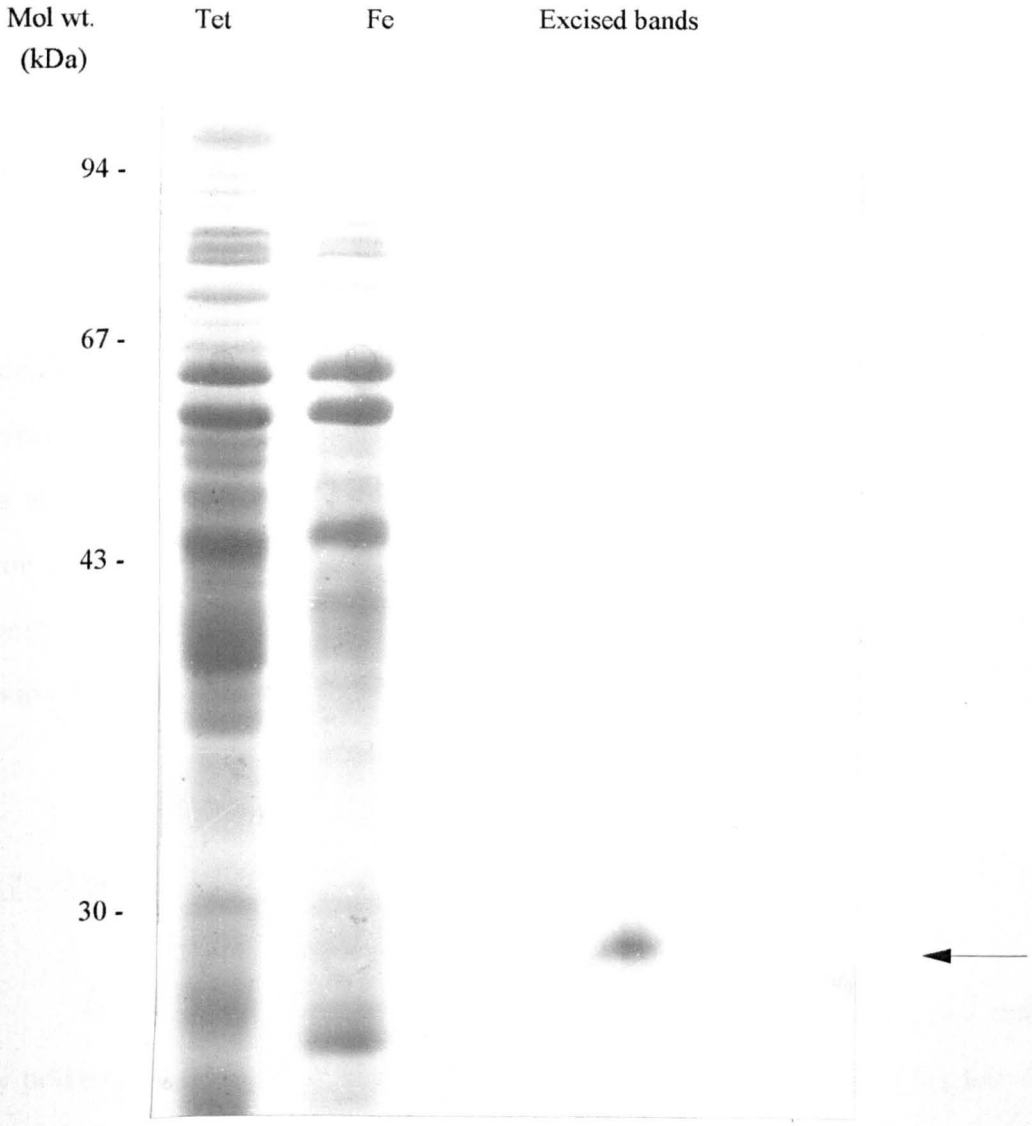
Native PAGE was carried out using iron-grown and tetrathionate-grown samples of *Sulfolobus* LM as above. Gels were examined for the presence of increased proportions of protein bands in the samples grown using ferrous iron.

Figure 5.3 Native PAGE of lysates of *Sulfolobus* LM grown using tetrathionate (Tet) and using ferrous iron (Fe), with high molecular weight protein standards (HMW std.). The band in excess in iron-grown cells is indicated.



The native polyacrylamide gel (Fig. 5.3) showed one protein band induced during growth on ferrous iron. It ran to the position of the 440 kDa standard on a native gel. The excision of this protein band from the native gel, followed by SDS treatment, allowed SDS-PAGE to be carried out using this protein.

Figure 5.4 SDS-PAGE of the protein from iron-grown cells which was eluted from a native polyacrylamide gel (Excised bands) compared with iron-grown (Fe) and tetrathionate-grown (Tet) cells. This band was induced during growth on ferrous iron.



This SDS polyacrylamide gel (Fig. 5.4) showed that the native protein giving a band at the position of the 440 kDa marker was comprised of 27 kDa polypeptides, thus explaining the band seen by single step SDS-PAGE of the iron-grown cell lysate. Elution of the protein from the gel slice caused it to run slightly behind the whole cell samples, as was confirmed by excising a 27 kDa polypeptide band from an SDS polyacrylamide gel and repeating SDS-PAGE. This showed the same behaviour as the polypeptide derived from the native band.

5.6 - Large scale autotrophic growth of Sulfolobus LM

For further protein characterisation to proceed, partial protein purification was necessary. This required relatively large amounts of the protein under investigation. *Sulfolobus* LM cells grown autotrophically on ferrous iron provided a very poor yield of biomass. A range of different growth strategies was considered. Previous work had shown that *Sulfolobus* LM failed to grow on yeast extract (Norris pers. comm.).

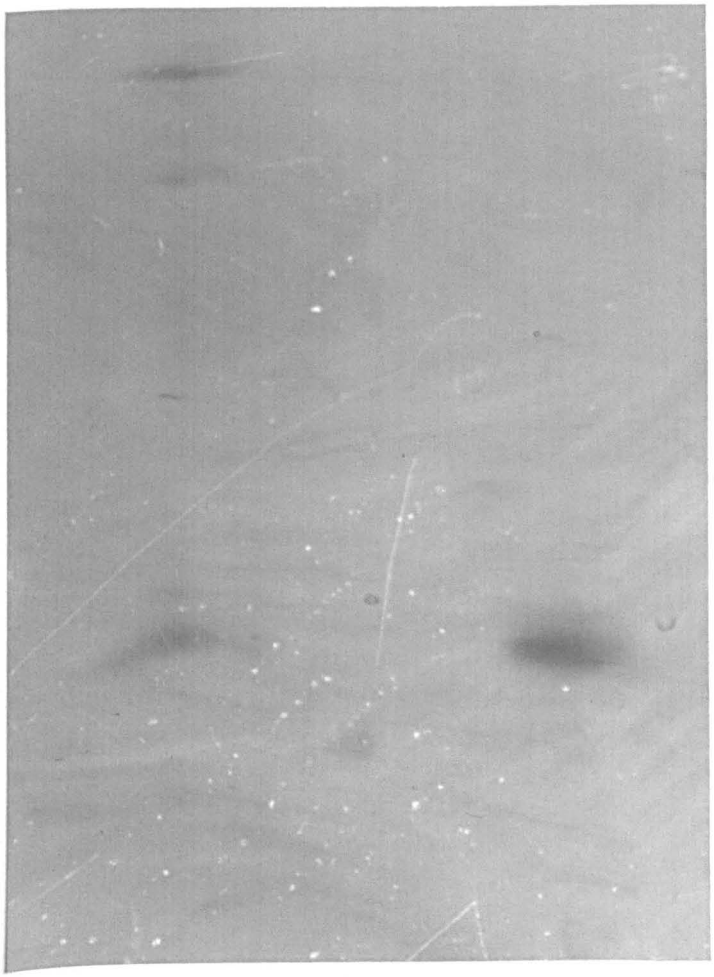
As detailed in section 6.4, induction of proteins related to iron oxidation by addition of ferrous iron to tetrathionate-grown and sulphur-grown cells was poor, and pyrite growth was problematical for a number of reasons (see section 6.4). Therefore, as alternative methods had proven unsatisfactory, autotrophic growth using ferrous iron as growth substrate was continued. 80 l batches of autotrophically grown cells were grown using four 20 l fermenters, cells being harvested after 60 % - 90 % oxidation of ferrous iron had occurred.

5.7 - N-terminal amino acid sequencing of 27 kDa protein

The partial purification of small quantities of the 27 kDa protein was achieved by polyacrylamide gel electrophoresis. Native polyacrylamide gels were carried out on 100 mg (wet weight) of these cells, yielding a large band of protein at the position of the 440 kDa marker. This was excised and subjected to the SDS sample preparation procedure, then used for SDS-PAGE. The SDS gel was subsequently blotted onto PVDF membrane which was then Coomassie stained (see Fig 5.5 overleaf). It was evident that this procedure had resulted in the separation of the 27 kDa subunit from the majority of contaminating polypeptides contained within the area of the native protein band.

Figure 5.5 PVDF membrane blot of SDS-PAGE of excised iron induced bands from a native gel (Native) and from an SDS gel (SDS). The position of the 27 kDa band is indicated.

Native SDS



- 27 kDa

Although the purification was probably incomplete, it was estimated that the 27 kDa band was present at many times the concentration of the second most abundant protein in the preparation. The area of membrane containing this band was sent for N-terminal amino acid microsequencing. Initially only six amino acids were identified from the N-terminus of this polypeptide. However, when the preparation was scaled up to using 500 mg (wet weight) of iron-grown cells, the sequence of twenty-five residues with three unknown amino acids was obtained.

Figure 5.6 N-terminal amino acid sequences of 27 kDa polypeptide

Initial sequence ;-

NH_2-Met-Lys-Leu-Tyr-???-Lys-Phe-

Subsequent sequence ;-

NH_2-Met-Lys-Leu-???-???-Lys-Phe-Pro-Glu-Thr-Gln-Val-Ile-Thr-Thr-Lys-Gly-Pro-Leu-Asp-Phe-Tyr-???-Asp-Val-Phe-Glu-

Composite sequence ;-

NH_2-Met-Lys-Leu-Tyr-???-Lys-Phe-Pro-Glu-Thr-Gln-Val-Ile-Thr-Thr-Lys-Gly-Pro-Leu-Asp-Phe-Tyr-???-Asp-Val-Phe-Glu-

5.8 - Design of DNA probes to the gene coding for the 27 kDa protein

The N-terminal sequence data was back translated using the genetic code and represented as follows ;-

NH_2-Met-Lys-Leu-Tyr-???-Lys-Phe-Pro-Glu-Thr-Gln-Val-Ile-Thr-Thr-Lys-

```
ATG AAG TTG TAT ??? AAG TTT CCA GAA TAT CAA GTA ATA ACA ACA AAG
      A   A   C       A   C   C   G   C   G   T   T   T   T   A
        C G                   T               G   C   G   G
        C A                   G               C       C   C
        C T
        C C
```

Gly-Pro-Leu-Asp-Phe-Tyr-???-Asp-Val-Phe-Glu-

```
GGA CCA TTG GAT TTT TAT ??? GAT GTA TTT GAA
  C   C   A   C   C   C       C   T   C   G
  G   T C G                       G
  T   G C A                       C
        C T
        C C
```

Initially, a probe was constructed to the portion of the N-terminal sequence following on from the second lysine residue. This portion avoided the wobble positions of leucine, valine and threonine which would have necessitated increasing the degeneracy of the DNA probe greatly, and avoided the unknown residues. A 20-mer oligonucleotide, suitable for probing and sequencing was designed using the codon bias of *Sulfolobus acidocaldarius* thermopsin (Lin *et al.*, 1990), *Sulfolobus acidocaldarius* sox terminal oxidase complex (Lubben et al., 1992), and *Sulfolobus solfataricus* aspartate aminotransferase (Cubellis *et al.*, 1989). These sequences all show a bias towards A and T in the wobble positions of proline, lysine and valine, as would be expected in these organisms of low GC content.

The antisense oligonucleotide was constructed, giving the sequence ;-

```
5'-ACT TGA GTT TCA GGA AAT TT-3'
     C   T   C   T   G   C
```

This probe was a 20-mer with 64 degeneracies, and a theoretical melting temperature of 59 ^{o}C in 6 x SSC, it was designated probe 27-1.

Later, an additional probe was constructed, this was made using the sequence from the first valine residue. Due to the greater number of possible wobble position uncertainties in this region, the codon bias was relied on to a greater degree than when designing the first probe. This probe was also produced as antisense sequence ;-

```
5'-GGA CCT TTA GTA GTA ATA AC-3'
     T   C   T   T   T   T
```

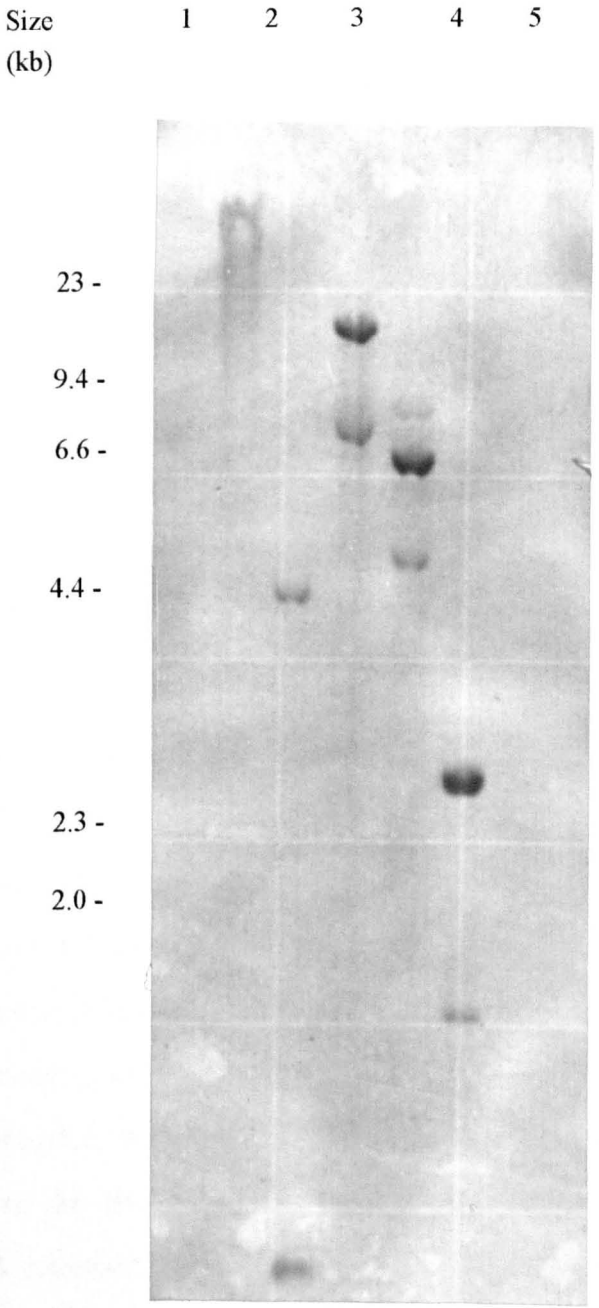
This second probe was a 20-mer with 64 degeneracies, and a theoretical melting temperature of 54 ^{o}C in 6 x SSC. It was designated 27-2.

5.9 - Detection of the gene coding for the 27 kDa polypeptide in *Sulfolobus* LM

Restriction digests of genomic DNA from *Sulfolobus* LM using *Eco* RI, *Bam* HI, *Hind* III and *Eco* RV were Southern blotted and probed using DIG labelled probe 27-1. Hybridisations were carried out at 59^{o}C with 5 x SSC and 2 x SSC washes. Bound oligonucleotide was detected colorimetrically (see Fig. 5.7 overleaf).

For each digest, save *Eco* RI, this showed a major band of oligonucleotide binding and a minor band. By reference back to standard DNA run on the original agarose gel these were found to have hybridised to genomic DNA fragments of approximately, for *Eco* RI 1 kb and 3.7 kb, for *Bam* HI 15 kb and 6.8 kb, for *Hind* III 6 kb and 4.5 kb, and for *Eco* RI 2.2 kb and 1.5 kb.

Figure 5.7 Southern blot of *Sulfolobus* LM genomic DNA, (1) undigested, and digested by (2) *Eco* RI, (3) *Bam* HI, (4) *Hind* III and (5) *Eco* RV and probed with oligonucleotide 27-1 at 59°C.



5.10 - Isolation of possible clones homologous to oligonucleotide 27-1

From the Southern blot (Fig. 5.7) it appeared that there were two genes to which probe 27-1 hybridised. Hybridisation of probe 27-2 was attempted in similar experiments, but no clear hybridisation was seen. This may have been due to insufficient labelling or incorrect oligonucleotide synthesis. Partial libraries of *Sulfolobus* LM genomic DNA fragments of between 6 and 7 kb and of between 4 and 5 kb were constructed. The fragments were formed by restriction with *Hind* III then eluted from an agarose gel.

A pBluescript KS+ phagemid vector was chosen as the cloning vector. This is a high copy number vector which has a large polylinker and ampicillin resistance and is amenable to blue / white selection of transformants with inserts. The vector was restricted with *Hind* III, phosphatased to prevent self-religation and repurified. After purification, the genomic DNA fragments were ligated into the vector, then transformed into *E. coli* strain TG 1. Sets of 500 white colonies were selected and replated in duplicate. Sets of colonies were colony blotted and hybridised with DIG labelled probe 27-1 at 45^{o}C with washes in 5x and 2x SSC at the same temperature.

Visualisation of colonies to which the probe had hybridised proved difficult with the DIG colorimetric detection system. However, two different clones from the 4 - 5 kb partial genomic library were isolated. These were designated 4.5-4 and 4.5-15. Both were sequenced upstream using the probe 27-1 as a sequencing primer, new primers were constructed, then both clones were sequenced downstream to reveal the DNA sequences at the binding sites of probe 27-1. Unfortunately, neither clone possessed a DNA sequence capable of producing the N-terminal sequence detailed above. One clone possessed 12 bases homologous to 27-1, the other possessed 15 bases homologous to 27-1. Therefore this cloning attempt had proven unsuccessful.

5.11 - Cloning and sequencing of the gene encoding the 27 kDa polypeptide

The cloning experiments detailed above were repeated by Dr. N. Burton using radioactively end-labelled oligonucleotide probes 27-1 and 27-2. These gave greater sensitivity and clearer results with colony hybridisations than the DIG labelled probes. This allowed the isolation of a 4.5 kb clone from a 3.5 - 5 kb *Hind* III partial genomic library of *Sulfolobus* LM. The sequencing of this clone was carried out by manual and automatic sequencing and revealed a DNA sequence corresponding to that predicted by the N-terminal amino acid sequence, all predicted amino acids were identical and the two unknown residues were revealed as glutamine and arginine. This sequence was situated at the beginning of an open reading frame encoding a polypeptide of 215 amino acids. The molecular mass of this polypeptide was calculated to be 25,015 Da. This figure is similar to that derived from SDS-PAGE of approximately 27 kDa.

Therefore, as the N-terminal sequences were identical over 25 amino acids and the molecular weight close to that derived from SDS-PAGE it was possible to claim with a high degree of confidence that this was the gene coding for the 27 kDa iron-induced polypeptide.

5.12 - DNA structure

There was no evidence of a leader sequence and the first codon was immediately preceded by a stop codon (see Fig. 5.8). No other significant open reading frames (>150 bp) were found downstream (270 bp sequenced) or upstream (300 bp sequenced) of the gene. A number of transcriptional control features were evident, but there was no Shine-Delgano sequence near the first codon. Two potential box-A sequences (AATATA and GTTATA) centred 60 and 75 bp upstream of the first codon

were similar to the promoter consensus sequence TTTA(A or T)A of *Sulfolobus* and other archaea (Reiter *et al.*, 1990). There was an 8 nucleotide inverted repeat AAAAATTG (CAATTTTT) between 27 and 55 bp upstream of the first codon and a number of polythymidine stretches in the AT-rich region (75 of 83 nucleotides) between 77 and 160 nucleotides after the stop codon. Transcription termination in *Sulfolobus* has been associated with polythymidine sequences (Reiter *et al.*, 1988).

Figure 5.8 Nucleotide and amino-acid sequence of 27 kDa polypeptide

```
                                                    'box A'         'box A'
                                          AATGAACTCGTAAATATAGTAACTCAAGTT
          I.R.                    I.R.
ATATGCAAAAATTGCACAAAACATTCAATTTTTAGGCGTGTTCTCTCCAAGTTATATTAA -1

ATGAAGCTATATCAAAAGTTCCCTGAAACCCAAGTGATAACTACAAAAGGTCCCCTCGAT  60

M  K  L  Y  Q  K  F  P  E  T  Q  V  I  T  T  K  G  P  L  D    20


TTTTATAGGGATGTATTTGAGAAAGGTAAGTGGCTATTCCTCTTCGCTCATCCAGCTGAC 120

F  Y  R  D  V  F  E  K  G  K  W  L  F  L  F  A  H  P  A  D    40


TTTACTCCAGTCTGCACGACGGAGTTTGTGGGATTCTCAAAGGTCTATGAAGAATTCAAG 180

F  T  P  V  C  T  T  E  F  V  G  F  S  K  V  Y  E  E  F  K    60


AGACTCAACGTGGAGTTAGTTGGGATGAGTGTAGATAGCATCTATTCCCACATAGAGTGG 240

R  L  N  V  E  L  V  G  M  S  V  D  S  I  Y  S  H  I  E  W    80


CTAAAGGATATCCAAGAGAGATATGGAATACAAGTTCCGTTTCCCTTAATAGCTGATCCA 300

L  K  D  I  Q  E  R  Y  G  I  Q  V  P  F  P  L  I  A  D  P   100


GATAAGAGGCTTGCTAGACTTCTTGATATAATTGATGAAGCCTCAGGAGTTACAATAAGA 360

D  K  R  L  A  R  L  L  D  I  I  D  E  A  S  G  V  T  I  R   120
```

GCAGTCTTTTTAGTTAACCCAGAAGGCATAATCAGATTTATGGCCTATTATCCTATCGAG 420

A V F L V N P E G I I R F M A Y Y P I E 140

TATGGCAGAAAAATAGAGGAGCTCTTAAGAATAACCAAGGCTGCCCTTGTAAACTATAAG 480

Y G R K I E E L L R T K A A L V N Y K I 160

GCAAAGGTTTCTCTACCAGTAGACTGGGAACCAGGTCAGGAGGTTATAGTTCCAGCTCCA 540

A K V S L P V D W E P G Q E V I V P A P 180

TCTACAATCGATGAAGCTCAGATAAGGATGAAGTTACCAAACGCTAAGACTTGGTATCTA 600

S T I D E A Q I R M K L P N A K T W Y L 200

ACCTTCAAGAAATATGATGAGTTACCTCAGGATCAAAGGGTAGTATAAACTTTAATAAAT 660

T F K K Y D E L P Q D Q R V V * 215

CTCTTTCAAATTGCTAAAGTAAGTTATATAGATAATAATAGAAAAACTATAGTATAAGGT 720

term ? term ?

TTAAAACATATATATAAAATAGAAAGAAAATTTTTTAATATTTAATTAAATAAAGTTTCT 780

term ?

ATAGAATTTTTTTAGGTAAAGATACAGAAGATATTTAGTATCTTGTAAAATACCTTATAG 840

CTAAAACGAGATTCAGTACCATAAATTTTGTAATTGTATAAACTCCTCGGC 891

5.13 - Protein homologies

The amino acid sequence of the *Sulfolobus* LM 27 kDa polypeptide was used for database homology searches by the BLAST program. The searches revealed homology with the alkyl hydroperoxide reductase / thiol-specific antioxidant (AhpC/TSA) family of proteins. Examples were compared with the sequence of the 27 kDa polypeptide. Alignments were represented diagramatically (Figure 5.9) and amino acid homologies were shown in full (Figure 5.10).

The AhpC/TSA proteins may be split into three groups. The *Sulfolobus* 27 kDa polpeptide is most similar to members of the first group, which includes examples from yeast and barley, and another example from the Archaea, the predicted product of a gene (*ORF K*) which is adjacent to the superoxide dismutase gene in *Methanobacterium thermoautotrophicum*. The proteins in this first group possess one conserved cysteine-containing region which has the sequence FTPVCTTE.

A second group contains the majority of the AhpC/TSA proteins described to date. Three eukaryotic examples and three bacterial examples are shown (Figure 5.10, middle group of six sequences). These have two conserved cysteine-containing regions, the first sequence, TFVCPTE, being completely conserved, while the second region is more variable. Some eukaryotes, for example yeast and human, have been shown to possess both 1 cysteine and 2 cysteine proteins.

A protein from *Plasmodium falciparum* (Fig. 5.10, bottom sequence) and a bacterioferritin co-migratory (BCP) protein from *Escherichia coli* (Andrews *et al.*, 1991) may belong to a third group, with a TPGCT sequence for the single, conserved cysteine-containing region. Six amino acid residues are conserved in all of the aligned proteins and other known members of the family.

The amino acid sequences of the proteins abbreviated in figures 5.9, 5.10 and 5.11 are of the following proteins ;-

Mthe - *Methanobacterium thermoautotrophicum* ORF k gene product (Meile *et al* ., 1995)

S27 - *Sulfolobus* LM 27 kDa polypeptide

Scer 1 - *Saccharomyces cerevisiae* ORF *YBL0524* product (Accession no. P34227)

Bsec - *Bromus secalinus* brome grass *pBS128* gene product (Goldmark *et al.*, 1992)

Hvul1 - *Hordeum vulgare B15C* barley gene product (Aalen *et al.*, 1994)

Hsap1 - *Homo sapiens HUMORF06* gene product (Accession no. D14662)

Scer2 - *Saccharomyces cerevisiae* thiol-specific antioxidant protein (Chae *et al.*, 1993)

Mmus2 - *Mus musculus* mouse peritoneal macrophage protein (Ishii *et al.*, 1993)

Hsap 2 - *Homo sapiens* TSA-antioxidant protein (Lim *et al.*, 1994)

Btau - *Bos taurus* bovine adrenal cortex protein (Watanabe *et al.*, 1994)

Mmus1 - *Mus musculus* mouse MER5 protein (Yamamoto *et al.*, 1989)

Cpas - *Clostridium pasteurianum* product of gene adjacent to rubredoxin gene (Mathieu *et al.*, 1992)

Ehis - *Entamoeba histolytica* peripheral membrane protein (Reed *et al.*, 1992)

Hvul2 - *Hordeum vulgare* gene product (Accession no. Z34917)

Hpy1 - *Helicobacter pylori* species-specific antigen (O'Toole *et al.*, 1991)

Eco1 - *Escherichia coli* AhpC protein (Accession no. P26427)

Styp - *Salmonella typhimurium* AhpC protein (Tartaglia *et al.*, 1990, Accession no. J05478 for corrected sequence)

Balc - *Bacillus alcalophilus* incomplete sequence of gene upstream of NADH dehydrogenase gene (Xu *et al.*, 1991)

Cdip - *Corynebacterium diphtheriae* iron-repressible gene product (Tai *et al.*, 1995)

Mavi - *Mycobacterium avium* antigen (Yamaguchi *et al.*, 1992)

Pfa1 - *Plasmodium falciparum* protein (Hudson-Taylor *et al.*, 1995)

Figure 5.9 Regions of homology between proteins of the AhpC/TSA family and the *Sulfolobus* 27 kDa polypeptide (after Ishii *et al.*, 1993). Black shading indicates highly conserved cysteine-containing regions, grey shading indicates regions of over 35 % identity with the *Sulfolobus* polypeptide sequence.

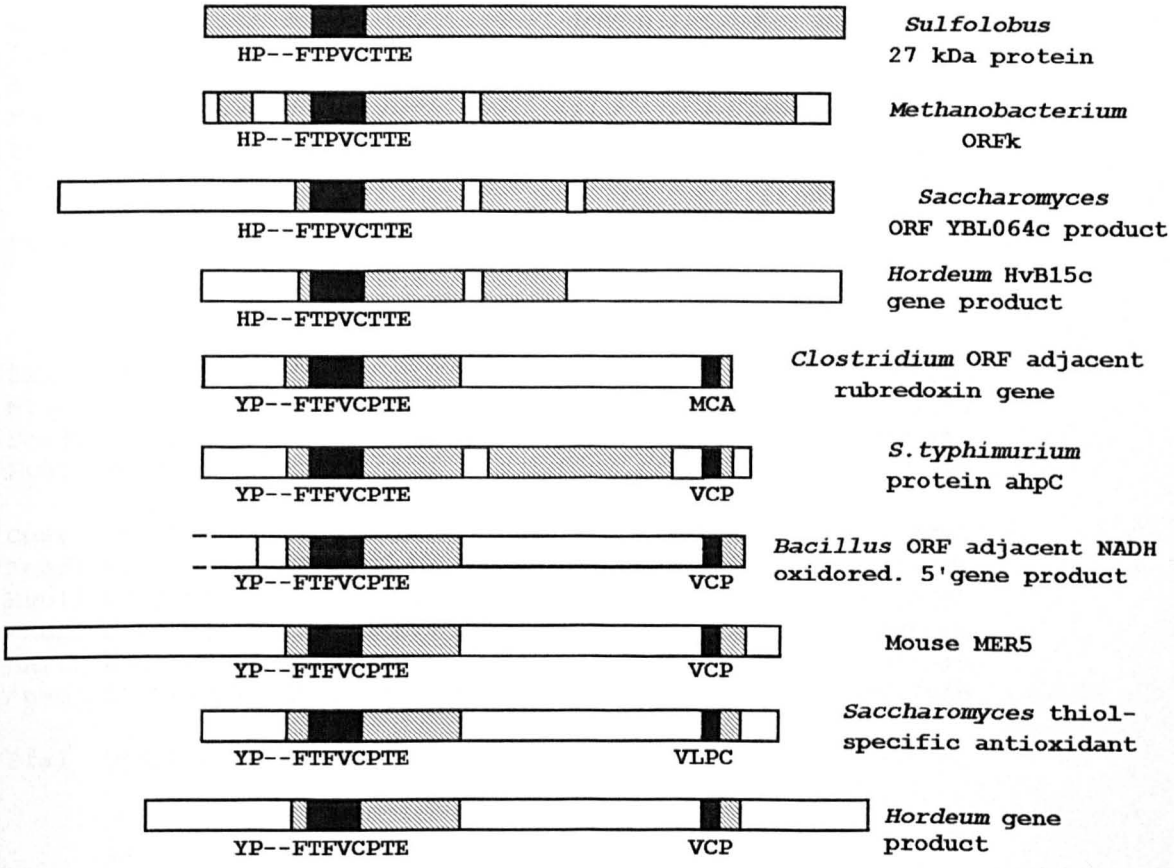


Figure 5.10 Alignment of the amino acid sequences of *Sulfolobus* strain LM 27 kDa iron-induced polypeptide and AhpC/TSA proteins.

```
      1                                                    cysteine 1
S27   MKLY.QKFPE TQVITTKGP. .LDFYRDVFE KGWKLFLFAH PADFTPVCTT
Mthe  MPLIGDKFPE MEVQTTHGL. .MKLPAEF.. KGRWFILFSH PADFTPVCTT
Scer1 RLRINSDAPN FDADTTVGK. .INFYDYL.. GDSWGVLFSH PADFTPVCTT
Hvul1 GLTIGDTVPN LELDSTHGK. .IRIHDYV.. GNGYVILFSH PGDFTPVCTT

Cpas  MKAVKGDGRG FT........ ..EVKLGDYK G.KWLVMFFT PLDFTFVCPT
Scer2 VAQVQKQAPT FKKTKVVDGV FDEVSLDKYK G.KYVVLAFI PLAFTFVCPT
Hvul2 LPLVGNKAPD FAAEKVFDQK FINVKLSDYI GKKYVILFFT PLDFTFVCPT
Mmus1 TPAVTQHAPY FKGTKVVNGK FKELSLDDFK G.KYLVLFFT PLDFTFVCPT
Balc  .......... .........K FIEVSEESFK G.QWSVLCFT PADFTFVCPT
Styp  MSLINTKIKP FKNQKFKNGK FIEVTEKDTE G.RWSVFFFT PADFTFVCPT

Pfa1  LDESIQKVEV LNHNGETTSF YNEVEKHKEN N.EGIVVFTY PKANTPGCTK
                                                   *   *  *

      48
S27   EFVGFSKVYE EFKRLNVELV GMSVDSIYSH IEWLKDIQE. RYGIQVPFPL
Mthe  EFVAFQEVYP ELRELDCELV GLSVDQVFSH IKWIEWIEE. NLDTEIEFPV
Scer1 EVSAFAKLKP EFDKRNVKLI GLSVEDVESH EKWIQDIKEI AKVKNVGFPI
Hvul1 ELAAMANYAK EFEKRGVKLL GISCDDVQSH KEWTKDIEAY KPGSKVTYPI

Cpas  EKTGFSKRAE EFRDLKAELL AWSCDSQYSH ETWINQDIKQ GGLGKTNFPI
Scer2 EKIAFSEAAK KFEEQGAQVL FASTDSEYSL LAWTNIPRKE GGLGPTNIPL
Hvul2 EKTAFSDRHE EFEKINTEIL GWSVDSVFSH LAWVQTERKS GGLGDLKYPL
Mmus1 EKVAFSDKAN EFHDVNCEVV AWSVDSHFSH LAWINTPRKN GGLGHMNITL
Balc  ELEDLQNEYA ALKELGVEVF SASTDTHFTH KGW...HDSS ETIGKTTYAM
Styp  ELGDVADHYE ELQKLGVDVY SWSTDTHFTH KAW...HSSS ETIAKTKYAM

Pfa1  QAELFKEKHE EFVNNKYVVY GLSADTAEDQ LKWKEKLE.. .....LPYEL
                               *          *

      97
S27   IADPDKRLAR LLDIIDE... .....ASGVT IRAVFLVNPE GIIRFMAYYP
Mthe  IADTG.RVAD TLGLIHP... ....ARPTNT VRAVFVVDPE GIIRAILYYP
Scer1 IGDTFRNVAF LYDMVDAEGF KNINDGSLKT VRSVFVIDPK KKIRLIFTYP
Hvul1 MADPDRSAIK QLNMVDPD.. .EKDAQGQLP SRTLHIVGPD KVVKLSFLYP

Cpas  ASDKTTEVVT KYGIQIEE.. ......EGIS ERGIFIKGPE GIVRYSVVHD
Scer2 LADTNHSLVR DYGVLIEE.. ......EGVA ERGIFIKGPK GVIRHITIGD
Hvul2 VSDVTKSIVK SFGVLIPD.. ......QGIA ERGIFIKGKE GVIQHSTIGN
Mmus1 LSDITKQIVR DYGVLLES.. ......AGIA ERGIFIKGPN GVVKHLSVGD
Balc  IGDPSQTLVR NFDVLNEV.. ......SGLA DRGTFIKGPD GVVQAAEIGA
Styp  IGDPTGALTR NFDNMRED.. ......EGLA DRATFVVGPQ GIIQAIEVTA

Pfa1  LCDVDKNLLK LLGLTNEE.. .......DKT IRSHLVLFKG DFTVSYVKKS
        *                                *
```

```
      139                    cysteine 2
S27   IEYGRKIEEL LRITKAAL.V NYKAKVSLPV DWEP....GQ EVIVPAPSTI
Mthe  QELGRNIPEI VRMIRAFR.V IDAEGVAAPA NWPDNQLIGD HVIVPPASDI
Scer1 STVGRNTSEV LRVIDALQ.L TDKEGVVTPI NWQP....AD DVIIPPSVSM
Hvul1 SCTGRNMDEV VRAVDSLL.T AAKHKVATPA NWKP....GE CVVIAPGVSD

Cpas  LNVGRSVDET LRVLKAFG.. ...TGGMCAL DWHE....GD DNL.......
Scer2 LPVGRNVDEA LRLVEAFGWT .DKNGTVLPC NWTP....GA ATIKPTVEDS
Hvul2 LGIGRSVDET LRTLQALGYV .KKPDKVCPA GWKP....GE KSMKPDLGPK
Mmus1 LPVGRSVEET LRLVKAFGFV .ETHGKVCPA NWTP....ES PTIKPSPTAS
Balc  EGIGRDASTL VNKIKAAGYV RNNPGKVCPA KWQE....GD ETLKPSLDLV
Styp  EGIGRDASDL LRKIKAAGYV AAHPGKVCPA KWKE....GE ATLAPSLDLV

Pfal  VSPGKSATQV LNFLVNGDDG GGDGNEEEEN EENNNNEDKD NNENDEEGDV
         *

      184
S27   DEAQIRMKLP NAKTWYLTFK KYDELPQDQR VV        from amino acid  1
Mthe  ETA..RKRKE EYECYDLALP .....PQRKW ..        from amino acid  1
Scer1 DEAKAKFGQ. .....FNEIK PYLRFTKSK. ..        from amino acid 48
Hvul1 EEAKKMFPQG FETADLPSKK GYLRFTKV.. ..        from amino acid  3

Cpas  .......... .......... .......... ..        from amino acid 14
Scer2 KEYFEAANK. .......... .......... ..        from amino acid  2
Hvul2 RSTRCYLERT FALSCGVLSW PFLGYMCFCD PS        from amino acid 17
Mmus1 KEYFEKVHQ. .......... .......... ..        from amino acid 63
Balc  GKI....... .......... .......... ..        from amino acid  1
Styp  GKI....... .......... .......... ..        from amino acid  1

Pfal  QGEGEGEGEG DEEKTADTDK EKPKKSSTST QK        from amino acid 10
```

The protein name abbreviations are given on page 139. Numbering is as for the *Sulfolobus* sequence. Residues identical to those of the *Sulfolobus* polypeptide are shown in bold print. The cysteine conserved regions are indicated. Residues conserved in all sequences are marked by an asterisk below the sequences.

5.14 - Comparisons between related proteins

Comparison of the predicted amino acid compositions of the above proteins, showed few deviations from a generally similar pattern apart from the example from *P. falciparum* and the *E. coli* BCP protein.

The amino acid composition of a range of these proteins was compared. In the case of the *Sulfolobus* polypeptide, only tyrosine (11 out of 215 amino acids) was noticeably more common than in most of the other proteins, including those of *M. thermoautotrophicum* (5/209), yeast (*YBL0524* gene product, 5/261; TSA protein, 5/195) and barley (*B15C* gene product, 7/218). The *Sulfolobus* polypeptide contained only one cysteine (at the conserved site) while there were three (one at the conserved site) in that of *M. thermoautotrophicum*, three (one at the conserved site) in the yeast *YBL0524* gene product, two (one at each conserved site) in the yeast TSA protein and four (one at a conserved site) in the barley *B15C* gene product. The *M. thermoautotrophicum* protein contained relatively few lysine residues, with 7 in comparison to 17 for the *Sulfolobus* polypeptide, 25 and 14 for yeast (*YBL0524* gene product and TSA protein respectively) and 20 for barley (*B15C* gene product). Arginine in the *M. thermoautotrophicum* protein conservatively replaced four of the lysines present in the *Sulfolobus* polypeptide.

5.15 - Evolutionary relationships between AhpC/TSA proteins

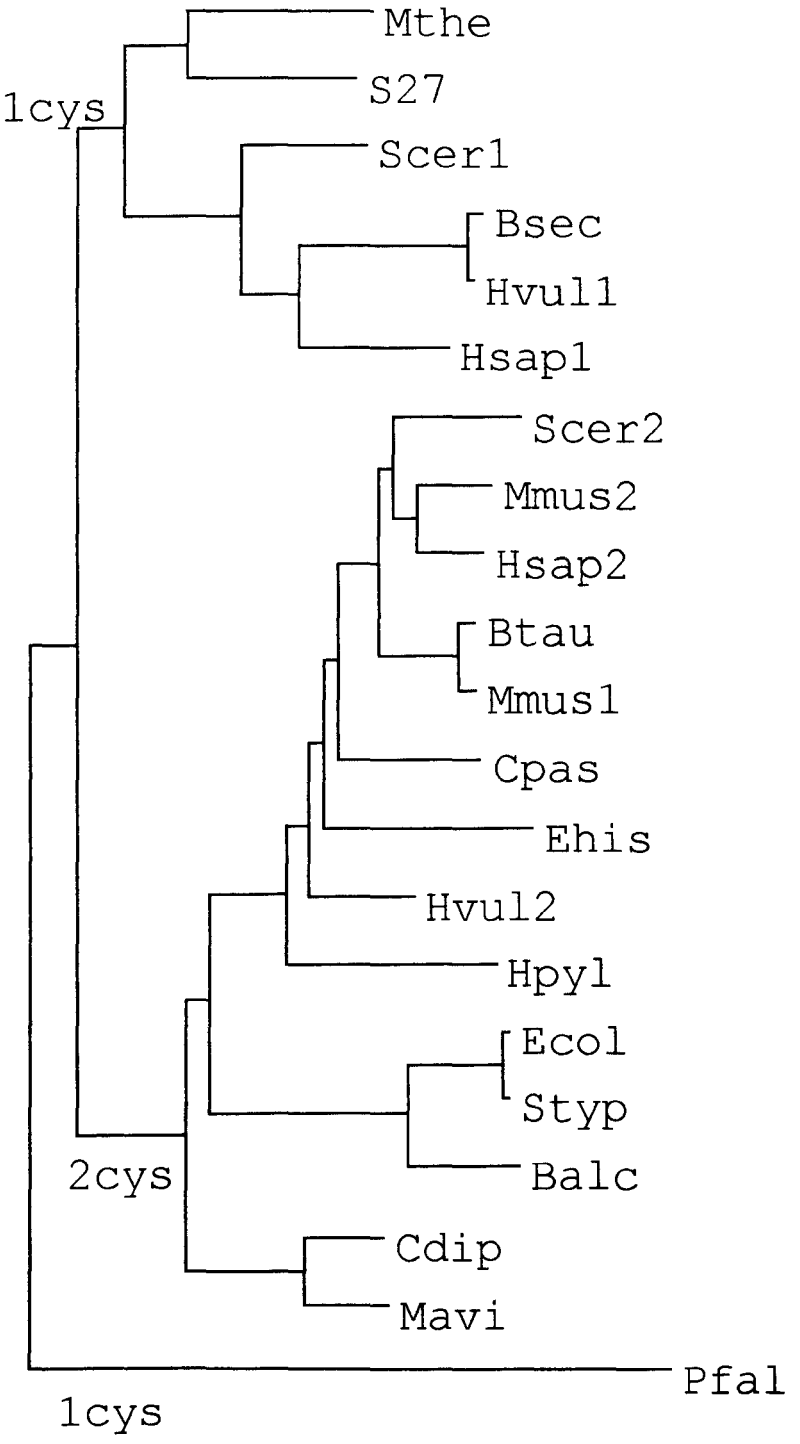
Amino acid sequences of twenty-one AhpC/TSA proteins were compared using GCG computer programs. Sequences were aligned and alignment was optimised. A similarity dendrogram was drawn for the aligned sequences (see Fig. 5.11) using the Fitch program from a distance matrix derived using the Protdist program of PHYLIP (Felsenstein, 1995).

This showed the sequence of the 27 kDa polypeptide clustering together with the only other archaeal sequence in the 1 cysteine subdivision, that of *Methanobacterium autotrophicum*. This subdivision also contained the 1 conserved cysteine proteins from human, barley, bromegrass and yeast. This group did not contain the 1 cysteine enzyme of *Plasmodium falciparum*, which shows little similarity to any of the other proteins in its C-terminal region, and appears to represent a different group.

The group containing most representatives was that of the 2 conserved cysteine proteins. This split into two subdivisions, one containing the eukaryotic proteins, the other containing the eubacterial examples. There was one exception to this organisation, the *Clostridium pasteurianum* protein, which clustered with the eukaryotes.

As no eubacterial sequences are found in the 1 cysteine group, this tree would tend to support the hypothesis that the archaea are more closely related to eukaryotes than to eubacteria. However, the discovery of a eubacterial protein of the 1-cysteine group or a 2-cysteine archaeal protein would render this invalid. These types of proteins may well be identified in the future, but if found to be absent would support the close relationship of archaea and eukaryotes.

Figure 5.11 Similarity dendrogram of AhpC/TSA proteins including the 27 kDa polypeptide of *Sulfolobus* LM. For abbreviations see p. 139.



In *Sulfolobus* LM the 27 kDa polypeptide was induced during iron oxidation. However, in *Acidianus brierleyi*, this polypeptide was found not to be over-expressed during iron oxidation (Norris pers. comm.), suggesting that this polypeptide was not essential for iron oxidation. *Sulfolobus* LM was grown using tetrathionate in the presence of 5 mM nickel, cadmium and copper, reducing growth rates by approximately 50 %. However, in these three cases the 27 kDa polypeptide was not over-expressed (Norris pers. comm.). Additionally, when *Sulfolobus* LM was grown using tetrathionate and 50 mM ferrous iron added, no induction was seen (Norris pers. comm.). These findings argue against the 27 kDa polypeptide being involved non-specifically in metal resistance.

The 27 kDa polypeptide was found by sequence comparisons to be a member of the AhpC/TSA group of proteins. The precise functions of many proteins of this group have not been fully elucidated, but a role in mitigating the effects of oxidative stress appears to be probable.

The *S. typhimurium* AhpC protein, together with the AhpF protein (a homologue of thioredoxin reductase), forms alkyl hydroperoxide reductase, an enzyme responsible for the removal of potentially damaging alkyl hydroperoxides produced by oxidative stress (Tartaglia *et al.*, 1990). It has been proposed that the AhpC protein is responsible for reducing hydroperoxide substrates to the corresponding alcohol while the AhpF component reduces AhpC with the concomitant oxidation of NADH or NADPH (Jacobson *et al.*, 1989).

Induction of AhpC/TSA protein synthesis has been related to various stimuli. These include cadmium, zinc, arsenite, H_2O_2 (Ishii *et al.*, 1993) and cell differentiation (Yamamoto *et al.*, 1989) in mice, and osmotic shock in *Staphylococcus aureus* (Armstrong-Buisseret *et al.*, 1995). Iron in the growth medium, raised O_2 levels, and various sulphur-containing compounds have induced the yeast TSA protein (Lim *et al.*,

1994). In yeast *in vitro*, this protein prevents oxidative damage caused by mixed function oxidation systems (e.g. Fe^{3+}/O_2/thiols) capable of generating reactive sulphur species but not those (e.g. Fe^{3+}/O_2/ascorbate) producing only reactive oxygen species (Kim *et al.*, 1988). The possibility that the protein could eliminate peroxides in the manner of the AhpC protein of *S. typhimurium* has not been ruled out however and a more general "reducing" function for the AhpC/TSA family of proteins has been proposed (Chae *et al.*, 1994a).

The reason for the induction of the 27 kDa polypeptide in ferrous iron grown *Sulfolobus* LM is not known. As in the case of the yeast TSA protein, it could be hypothesised that the 27 kDa polypeptide from *Sulfolobus* is responsible for protection from reactive sulphur species generated by the interaction of ferric ions, oxygen and thiols. These may be present in sufficient concentrations to induce the 27 kDa polypeptide during autotrophic growth on ferrous iron, which requires small quantities of tetrathionate for biosynthetic purposes and results in the production of upto 50 mM ferric iron. However, growth using tetrathionate in the presence of upto 15 mM ferric iron did not result in induction of the 27 kDa polypeptide. This may indicate that the 27 kDa polypeptide does not protect against deleterious sulphur radicals, or, alternatively that the production of sulphur radicals varies greatly with ferric iron concentrations of 15 - 50 mM. Additionally, tetrathionate may exhibit the property of regenerating such an anti-oxidant protein, as was observed with glutathione in the case of yeast TSA and glutathione reductase (Chae *et al.*, 1994a).

An alternative function to that of protection against sulphur radicals may be protection against hydroperoxides. Hydroperoxides may be formed by the superoxide dismutase catalysed reaction of superoxide anions or the uncatalysed reaction of hydroperoxide radicals. Superoxide anions could perhaps be overproduced during respiration with ferrous iron as the substrate. *Sulfolobus* strains have shown superoxide dismutase activity comparable to that of *E. coli*, but are more sensitive to

H_2O_2, lacking catalase activity, but showing peroxidase activity with *o*-phenylenediamine as a substrate (Grogan, 1989).

Additionally, the gene coding for the AhpC/TSA homologue from *Methanobacterium thermoautotrophicum* is adjacent to the superoxide dismutase gene (Meile et al., 1995), perhaps implying some functional interrelation. At least eight AhpC/TSA proteins (Chae *et al.*, 1994b) are encoded by genes in close proximity to genes encoding various oxidoreductases, including the NADH dehydrogenase in an alkaliphilic *Bacillus* species (Xu *et al.*, 1991), and the NADH oxidases in *Amphibacillus xylanus* (Niimura et al., 1993) and *Streptococcus mutans* (D21803). There is also an ORF coding for a protein resembling a subunit of dimethylsulfoxide reductase adjacent to the BCP gene in *E. coli* (Andrews *et al.*, 1991). Genes adjacent to the 27 kDa polypeptide-encoding gene in *Sulfolobus* strain LM are currently unknown.

Predicting the exact function of the 27 kDa polypeptide from the primary amino acid sequence does not appear possible. It is, however, likely that homologous proteins possess similar functions in regions of high primary sequence similarity and secondary structure similarity. The secondary structures of both the *Methanobacterium thermoautotrophicum* protein and the *Sulfolobus* polypeptide are predicted, by the use of Chou-Fasman plots, to show little similarity to one-another or to other members of the AhpC/TSA family of proteins (Norris, pers.comm.).

Two signs of adaptation of the *Sulfolobus* polypeptide to activity at 70°C have been noted. In contrast to the mesophilic proteins the *Sulfolobus* polypeptide possesses no cysteine residues apart from the one at the conserved site, this may indicate increased protein stability as cysteine residues are potentially chemically modifiable. Also, an increase in the number of hydrophobic tyrosine residues may provide the

cohesive hydrophobic molecular interior thought to be necessary for protein thermostability (Hensel, 1993).

Further investigation of this protein by induction studies and, if possible, the study of non-producing mutants, may discover the true function of this protein in iron-oxidising *Sulfolobus* LM. If it is shown to be produced in response to oxidative stress, then it remains to be explained why it is not induced in *Acidianus brierleyi* grown under similar conditions.

CHAPTER 6

Investigation of a novel cytochrome from iron-oxidising *Sulfolobus* strain LM

6.1 - Introduction

The mineral sulphide-oxidising archaea oxidize sulphur compounds and ferrous iron, with a number of species being capable of autotrophic growth using a minimal salt medium, ferrous iron and a source of reduced sulphur for biosynthesis as they appear unable to use sulphate for this purpose (Marsh *et al.*, 1983). In contrast to well studied iron-oxidising eubacteria such as *Thiobacillus ferrooxidans* and *Leptospirillum ferrooxidans*, no abundant acid stable soluble electron carrier proteins are found in *Sulfolobus* LM during iron oxidation (Blake *et al.*, 1993).

However, the presence of a membrane-associated acid-stable cytochrome induced during iron oxidation has been described for *Sulfolobus* species HT and LM, *Metallosphaera sedula* and *Acidianus brierleyi* (Barr *et al.*, 1990). The further characterisation of this cytochrome required its isolation, a procedure which posed some difficult problems since extremely little biomass is produced during growth on ferrous iron. About 300-400 mg cell dry weight per mole of ferrous iron oxidized is the maximum biomass theoretically expected from iron oxidation (Ingledew, 1982), in practice a lower yield was achieved. For this work *Sulfolobus* LM was the species chosen.

6.2 - Autotrophic growth of *Sulfolobus* LM

Cells had to be grown autotrophically on ferrous iron and on tetrathionate, so that the presence of the iron-induced cytochrome could be visualised by spectrophotometry and as a protein by electrophoretic comparison of whole cell lysates.

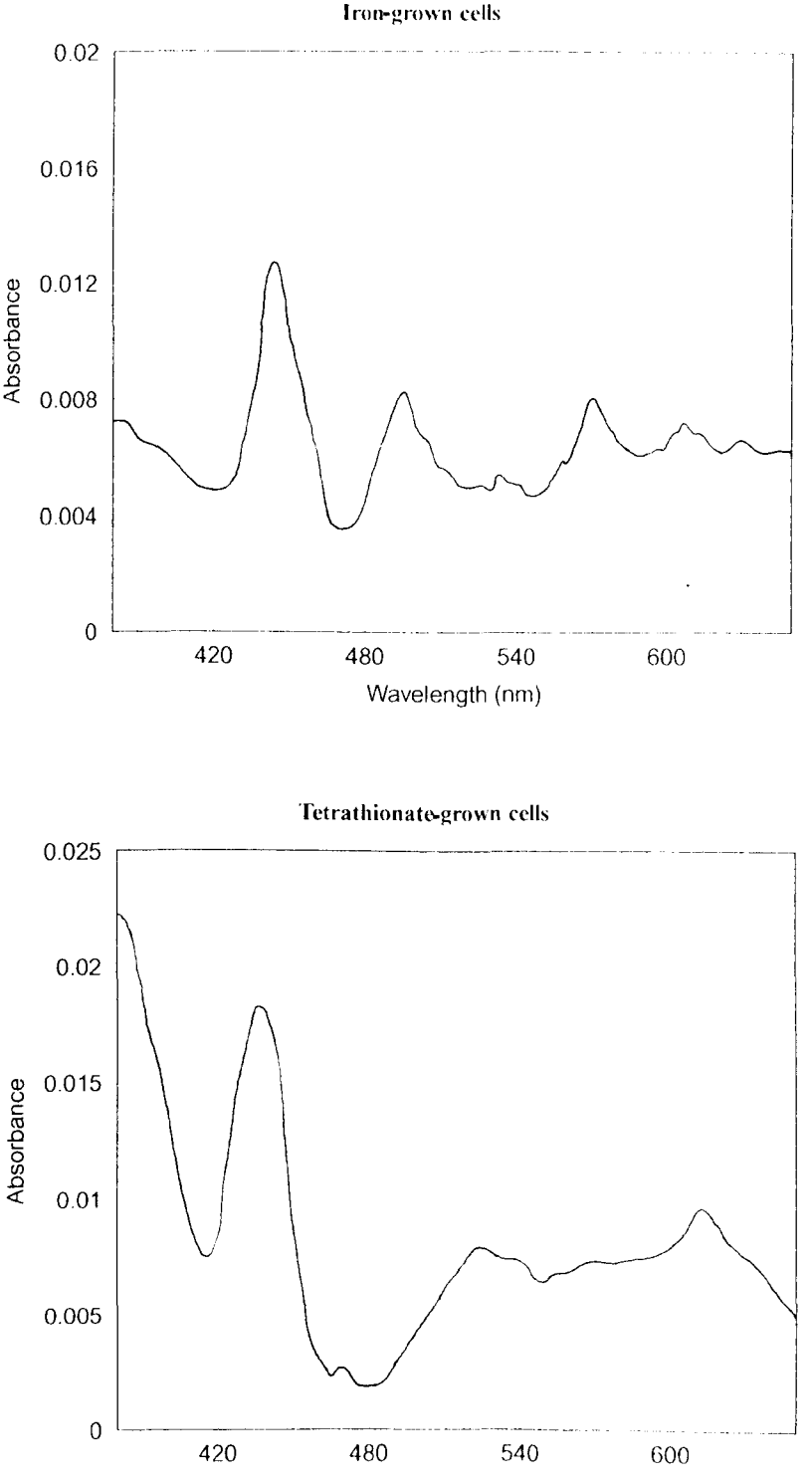
Sulfolobus LM was obtained from stocks grown using sulphur. Ten serial subcultures were carried out with cells grown on tetrathionate medium, and ten with cells grown on ferrous iron medium. This large number of serial subcultures was necessary in order to remove all traces of elemental sulphur from the medium, and to fully adapt cells to the substrates used, eliminating any proteins reflecting the past nutritional history of the cultures.

6.3 - Cytochrome identification by spectrophotometry

Optical spectrophotometry has been carried out on various iron-oxidising archaea, including *Sulfolobus* LM, *Acidianus brierleyi* and *Metallosphaera sedula*, showing absorbance peaks at approximately 483, 535, 572 and 603 nm, with a Sorêt peak at 440 nm (Barr *et al.*, 1990). The 603 nm peak has been attributed to the terminal oxidase aa_3-type cytochrome, which has been isolated and sequenced (Anemüller and Schäfer, 1990; Lübben *et al*, 1992), and is not unique to iron oxidation. However, the 572 nm peak appeared to be specific to ferrous iron oxidising cells.

Difference spectra were obtained using whole cells grown with ferrous iron and with tetrathionate (see Fig 6.1 overleaf). These showed the presence of a 572 nm peak in spectra of iron-grown cells but not in tetrathionate-grown cells thus confirming the results of earlier studies. A large Sorêt peak was present at 440 nm.

Figure 6.1 Difference spectra of oxidised-minus-reduced whole cells of *Sulfolobus* LM grown using iron and using tetrathionate.



6.4 - Large scale autotrophic growth of Sulfolobus LM

For further protein characterisation to proceed, protein purification was necessary. This required relatively large amounts of the protein under investigation. *Sulfolobus* LM cells grown autotrophically on ferrous iron provided a very poor yield of biomass. However, only cells grown with iron appeared to express the protein. Therefore, different growth strategies were considered to boost the biomass.

Pyrite (FeS_2) appeared to be a promising growth substrate, as it contains iron to induce the desired cytochrome and sulphur as an additional and energetically favourable substrate. However, being a natural mineral, it can contain a range of impurities. The available mineral, for example, was arsenopyrite; pyrite with a high arsenic concentration, which could have affected cell growth. Scale-up to 20 l batch fermenters was attempted but agitation of pyrite proved inefficient with available equipment, potential corrosion precluding the use of standard reactors. In addition, the 572 nm peak appeared to be induced to a lesser extent than when the cells were grown using ferrous iron as the substrate.

As problems were encountered with pyrite, *Sulfolobus* LM was grown on sulphur, followed by attempts to induce iron oxidation by the addition of ferrous iron when sufficient cell density was reached. *Sulfolobus* LM grew well on sulphur, and iron-induced cells were examined. However, no induction of the 572 nm peak could be detected, probably due to the excess sulphur still present in the medium being oxidized in preference to the iron.

Sulfolobus LM growth on tetrathionate is faster than on sulphur, and complete oxidation of the soluble sulphur compound is easier to obtain and monitor than when elemental sulphur is the substrate. Therefore, induction with ferrous iron was tried using tetrathionate as the initial growth substrate. The ferrous iron was added just

before cells reached the stationary phase, but again little induction of the 572 nm peak was noted.

These preliminary attempts to obtain larger quantities of iron induced biomass indicated that it was not a straightforward exercise and, if possible at all, would require considerable optimisation. Therefore autotrophic growth using ferrous iron as growth substrate was continued. In order to provide the gram quantities of biomass required for protein purification attempts, 400 l and 600 l of culture were grown in repeated runs with 20 l fermenters, cells being harvested after 60 % - 90 % oxidation of ferrous iron had occurred.

6.5 - *Sulfolobus* strain LM membrane preparation and solublisation

A protein band corresponding to the cytochrome which produces the 572 nm absorbance peak in iron-grown cells had not been identified previously by electrophoresis, the only iron-induced protein reproducibly visualised by SDS-PAGE was the 27 kDa polypeptide discussed in the previous chapter. Some purification of the cytochrome was intended to facilitate its identification by electrophoresis. Once it had been identified, it was intended to follow a similar route to characterisation as was used for the carboxylase and the 27 kDa iron-induced polypeptide. Additionally, further biochemical characterisation of the protein would be possible with purified or semi-purified cytochrome

The iron-induced cytochrome was thought to be membrane associated (Barr et al., 1990). Therefore the preparation of a membrane fraction from iron-grown cells was the obvious first step in purification. In addition, the 603 nm peak signifying the presence of cytochrome aa_3 as studied in heterotrophically-grown *Sulfolobus acidocaldarius* had been found in autotrophically-grown *Sulfolobus* LM. This protein

is known to be membrane associated (Anemüller and Schäfer, 1990) and was expected to provide a useful guide to progress in membrane purification. All procedures were tested using tetrathionate-grown *Sulfolobus* LM which expresses the 603 nm cytochrome but not the 572 nm cytochrome, but provides considerably more biomass per l than the ferrous iron substrate. This was to save the relatively limited amount of iron-grown biomass for use with optimised procedures.

Cells, typically 0.5 g wet weight, were lysed and membrane fractions prepared. Test preparations were carried out at pH 2 and at pH 7. Samples from various stages of the preparations were examined by redox spectrophotometry for the presence of active cytochromes. The presence of cytochromes was indicated by a Sorêt peak at 440 nm and by their alpha peaks at 572 or 603 nm. However, due to the low cytochrome concentrations in many preparations, the smaller alpha peaks were often not easily visible.

Initially, the ultracentrifuge cell fractionation procedure was tested at pH 2 and pH 7 with cells grown using tetrathionate. This procedure gave various fractions designated as follows ;-

Cells - washed whole cells of *Sulfolobus* LM, at pH 2 and 7

P1 - pellet 1, produced by 10,000 g centrifugation after cell lysis by French press

S1 - supernatant 1, containing both soluble and membrane fractions

P2 - pellet 2, produced by ultracentrifugation of S1, contains membrane fraction

S2 - supernatant 2, produced by ultracentrifugation of S1, contains soluble fraction

Redox spectra were carried out on these fractions. S1 samples at pH 7 and whole cells gave spectra showing both Sorêt and 603 nm alpha peaks clearly (see Fig. 6.2), the S1 sample at pH 2 showed an alpha peak of lower magnitude. S2 contained no cytochrome activity, whereas P2, when resuspended, showed only a trace of

cytochrome activity. This was probably due to the insoluble nature of the fraction. This experiment showed that the 603 nm cytochrome is probably membrane associated and had more activity at pH 7 than pH 2, although some activity was still retained at the lower pH. This agrees with the characteristics of the aa_3 terminal oxidase cytochrome (Anemüller and Schäfer, 1990).

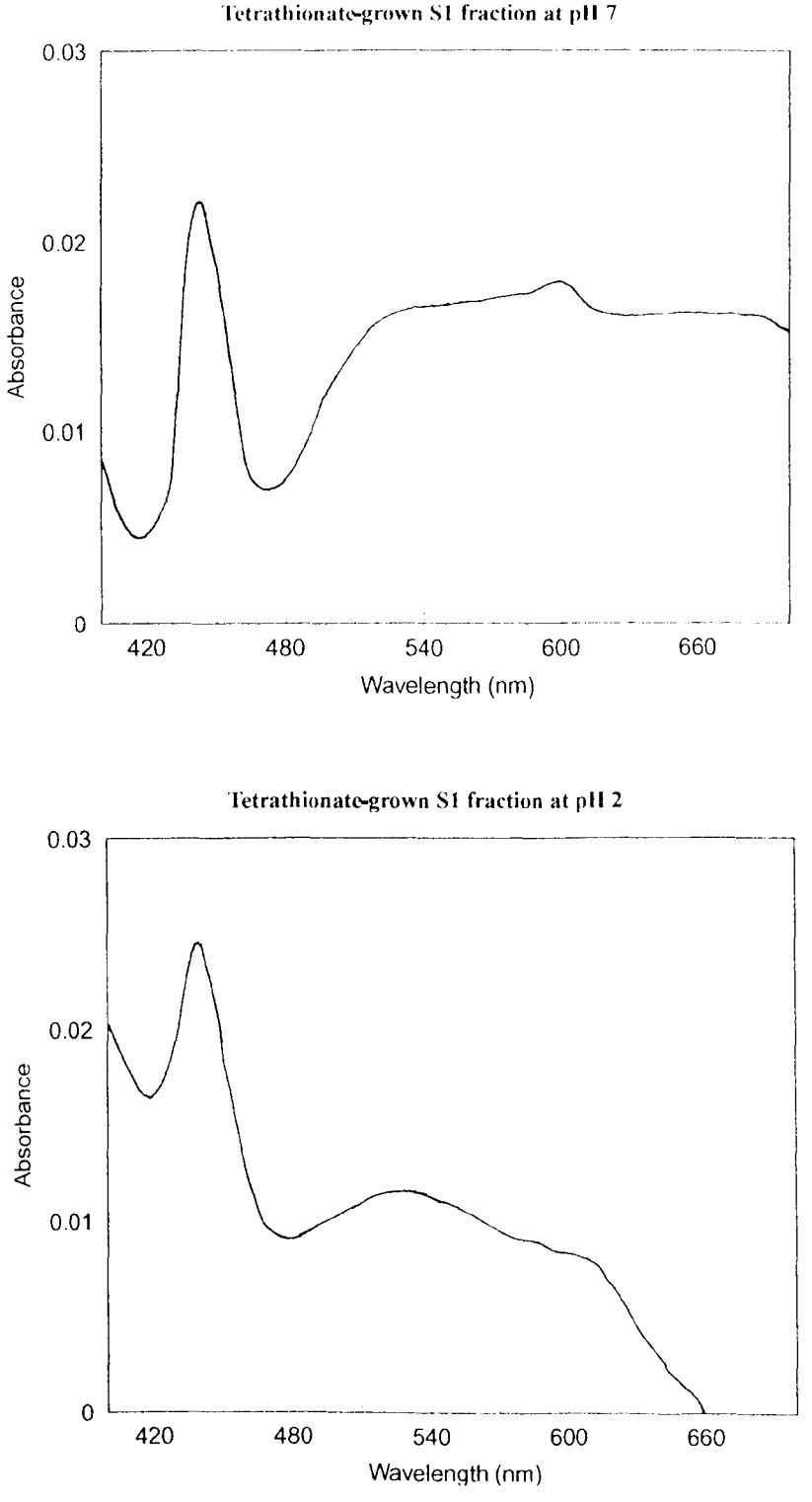
To confirm the presence of the 603 nm cytochrome in the membrane fraction and allow further investigation, the solublisation of the membrane pellet P2 was essential. Three different solublisation procedures were compared, using the detergents NP 40, MEGA 10 and sarkosyl. These preparations resulted in the production of two further fractions;-

P3 - Insoluble material, derived from solublisation and ultracentrifugation of P2

S3 - Solublised membrane components, derived from P2 as above

Redox peaks indicating cytochrome were found in S3, confirming the presence of cytochrome in the membrane fraction (see Fig. 6.3). Additional cytochrome may have been present in the insoluble P3, but this was not amenable to spectrophotometry. For a similar protein concentration, the NP 40 procedure resulted in double the Sorêt peak height (0.003 units) of S3 when compared with the other two methods, MEGA 10 (0.0012 units) and sarkosyl (0.0015 units). Therefore the NP 40 method, using a buffer designated SB II, was used for further preparations.

Figure 6.2 Difference spectra of S1 soluble fractions of tetrathionate-grown cells prepared at pH 2 and pH 7.



The membrane preparation and solublisation procedure was then used with iron-grown cells at pH 2 and at pH 7. It was noted that the lysis of cells was more efficient and easier at pH 7 than at pH 2, as expected for an acidophile. The membrane pellet P2 contained 1.25 mg protein when prepared at pH 2, as against 6 mg protein at pH 7, for a similar initial quantity of cells. The presence of the 572 nm peak and its accompanying 440 nm Sorêt peak was noted in whole cells and fractions S1 and S3, as predicted from the fractionation of the tetrathionate-grown cells. A 603 nm peak was also just visible. The 572 nm and Sorêt peak heights were greater in the pH 7 fractions than the pH 2 fractions, but allowing for the difference in total protein concentrations resulted in similar specific peak heights. For the S1 samples these were as follows ;-

572 nm at pH 2 = 0.004 units mg^{-1}

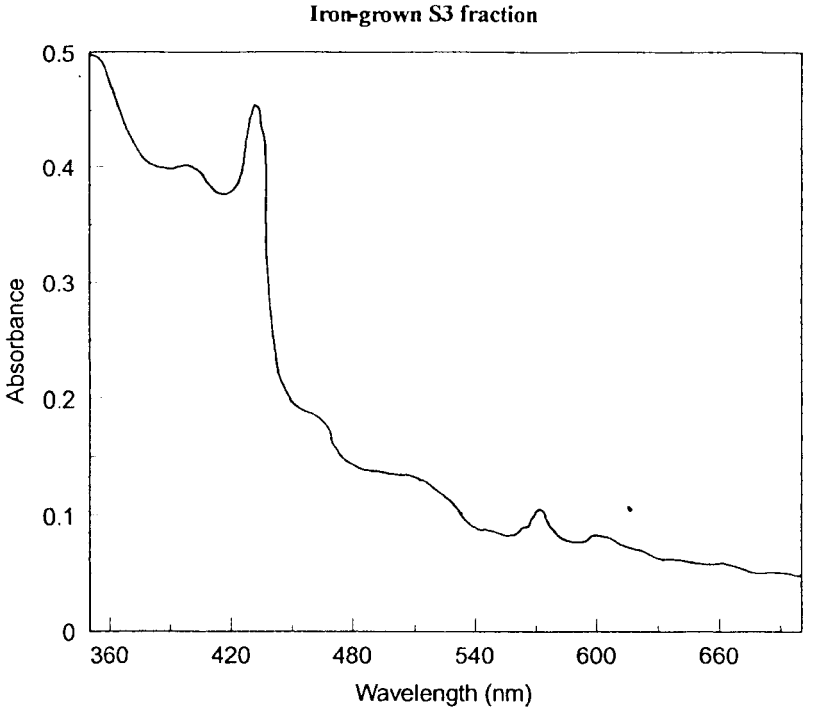
572 nm at pH 7 = 0.005 units mg^{-1}

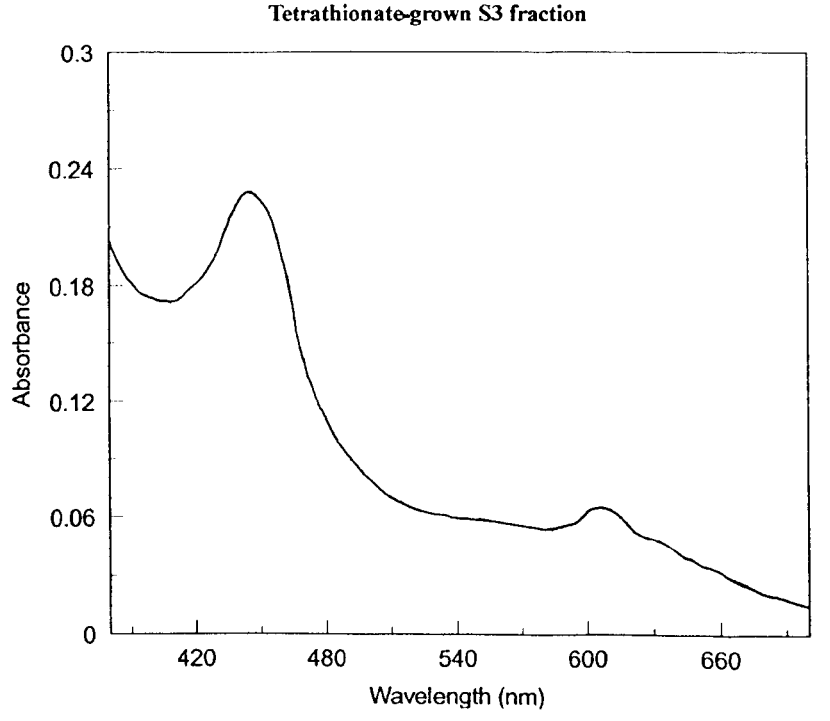
440 nm at pH 2 = 0.048 units mg^{-1}

440 nm at pH 7 = 0.055 units mg^{-1}

Therefore the 572 nm cytochrome was shown to be membrane associated and to be redox active at both pH 2 and at pH 7. It would be expected that a membrane protein exposed to the outside surface of the cell would be active at pH 2, whereas one exposed to the cytoplasm would be active at pH 7, close to the internal pH of *Sulfolobus* of 6.5 (Blake et al., 1993). The above results, however, would favour either location, or a protein capable of activity in both environments. The S1 and S3 fractions both had a yellow / green appearance, perhaps suggestive of an iron-containing cytochrome. The following figure (Fig. 6.3) illustrates the difference in spectra between the S3 solublised membrane fractions of iron- and tetrathionate-grown cells. This shows the relative amounts of 572 nm and 603 nm peaks. As well as redox spectra, spectra of samples versus buffer were carried out. This resulted in the shifting of the Sorêt peak from 440 nm to 420 nm. Therefore non-destructive testing of samples for the presence of cytochromes was shown to be possible.

Figure 6.3 Difference spectra of S3 solublised membrane fractions of iron-grown and tetrathionate-grown cells.





In addition to the procedure described above, the method originally used for the purification of cytochrome aa_3 (603 nm) from heterotrophically-grown *Sulfolobus acidocaldarius* (Anemüller and Schäfer, 1989) was tested using tetrathionate-grown cells of *Sulfolobus* LM. In this case it resulted in the loss of 75 % of cytochrome activity as assessed by Sorêt peak heights. This method was therefore considered unsuitable for further use with the limited quantity of autotrophically grown biomass.

In membrane preparations from iron-grown cells, the presence of an oily green layer between the pellet of insoluble material and the supernatant was usually seen. A similar observation had been made with *Sulfolobus acidocaldarius*, where this layer contained the majority of membrane bound cytochromes (Anemüller and Schäfer, 1989). Initially this layer had not been observed due to the small pellet size obtained from early preparations, and had probably been included in supernatant S3 fractions. On carrying out spectrophotometry, it was observed that of the three final fractions, S3, oily and P3, the oily layer and insoluble layer contained the greatest total and specific cytochrome activity. Figure 6.4 shows the redox spectrum of the oily layer.

572 nm peak in whole cells = 0.60 units for specific activity of 0.007 units mg^{-1}

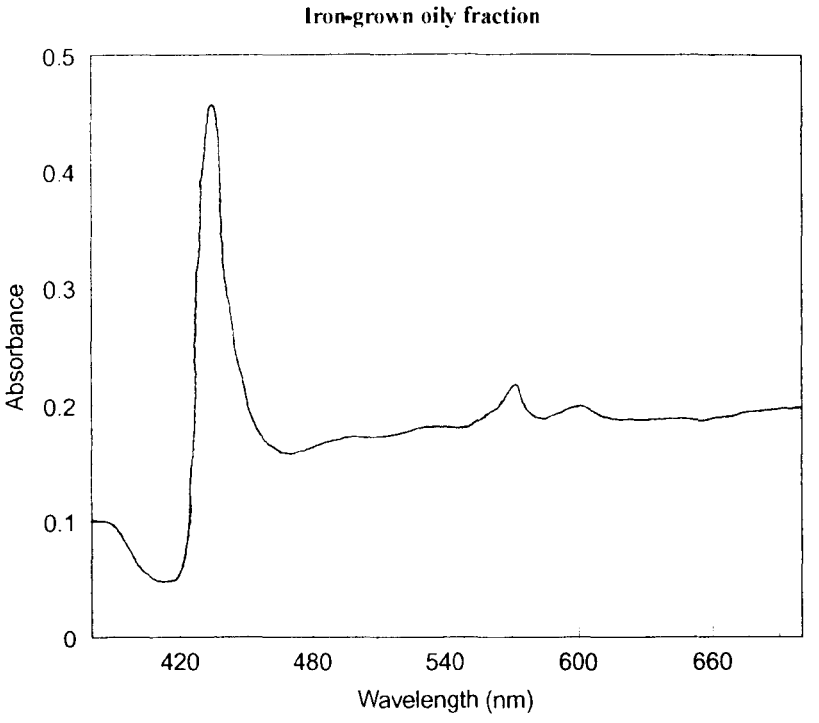
572 nm peak in S3 = 0.08 units for specific activity of 0.006 units mg^{-1}

572 nm peak in oily layer = 0.21 units for specific activity of 0.040 units mg^{-1}

572 nm peak in P3 = 0.20 units for specific activity of 0.040 units mg^{-1}

These results showed a purification factor for the 572 nm peak in the oily layer of 6 fold. However, they also showed that the solublisation procedure was relatively inefficient. The greatest solublisation of cytochrome was achieved by treatment of the pellet P3 with the chaotropes KSCN and DMSO and sonication of the sample prior to ultracentrifugation. This resulted in the solublisation of approximately half the cytochrome activity remaining in the pellet P3.

Figure 6.4 Difference spectrum of the oily layer derived from fractionation of *Sulfolobus* LM autotrophically grown using ferrous iron.



Samples showing the 572 nm peak were compared with those showing no such cytochrome activity by SDS-PAGE in the hope of identifying a protein band corresponding to the 572 nm cytochrome. No reproducible differences were seen. A TMBZ haem stain (Goodhew *et al.*, 1986) was carried out on a native polyacrylamide gel with samples showing cytochrome activity. This showed haem-stained material retarded at the interface between the stacking gel and the resolving gel. However, when this material was excised and used for SDS-PAGE, it was apparent that this band was comprised of a large number of separate proteins, providing no opportunity for the identification of the haem-staining protein. The use of a native polyacrylamide gel with additional NP 40 detergent to maintain solublisation of the sample in the resolving gel was attempted, but a similar result to the native polyacrylamide gel was found. Therefore further purification of the solublised cytochrome was deemed necessary to facilitate the visualisation of such a protein band.

6.6 - Cytochrome purifications by column chromatography

A number of different purification procedures were attempted using solublised membrane fractions S3 of tetrathionate- and iron-grown cells and the oily layer derived from iron-grown cells. Fractions were initially assessed for cytochrome activity by reading the absorbance of their Sorêt peaks at 420 and 440 nm versus the relevant buffer. This was found to be unreliable as the peak heights above the variable background absorbance were found to provide more indication of cytochrome activity. Therefore all samples were tested by scanning spectrophotometry and peak heights above background were calculated.

DEAE Sephacel is a weak anion exchange matrix, this was used in purification attempts using S3 from iron-grown cells and S3 from tetrathionate-grown cells. In the case of iron-grown S3, the eluted fractions all possessed proportional protein

concentrations and cytochrome concentrations, indicating the presence of no factor of purification. Fractions from tetrathionate-grown cells simply lost cytochrome activity.

Q-Sepharose FPLC was carried out on iron-grown S3. Elution of cytochrome activity was seen at a concentration of 150 mM NaCl. However, most cytochrome activity was lost. SDS-PAGE was carried out to compare fractions with and without cytochrome activity but no protein bands corresponding to cytochrome peak heights were distinguished. Positive fractions from this purification attempt were pooled and concentrated and used for size exclusion chromatography with a Superose 12 column by FPLC. The activity eluted was again much reduced, but some fractions were seen to contain a trace of cytochrome activity. These were again compared with negative fractions by SDS-PAGE, but again no protein bands corresponding to cytochrome peak heights were distinguished.

A Mono Q HR 5/5 anion binding FPLC column was used for a purification attempt with tetrathionate-grown S3. This resulted in the loss of 80 % of cytochrome activity, so was not used with iron-grown material.

Therefore no column purification attempts had resulted in satisfactory purification or identification of the 572 nm cytochrome. The problems with loss of activity may have been caused by sample contamination, protein denaturation by salt, an excessively avid binding to the columns, or simple adsorption of the small quantities of sample available to the apparatus. Due to the lack of available material, full investigation by repeated testing and alteration of chromatographic conditions was not possible. Therefore, a different approach was taken whereby very small quantities of samples were tested for binding to 1 ml conventional chromatography minicolumns.

Size exclusion matrices Sephacryl S200 and Sepharose CL6B were used for purification attempts on S3 fraction of tetrathionate-grown cells. However, in both

cases the cytochrome was eluted in the void volume indicating no retardation by the column. This suggested that the cytochrome may have still been associated with other membrane components such as lipids or with detergents. Micelles would have been too large to bind to size exclusion matrices.

DEAE Sephacel and DEAE cellulose were also used for purification attempts on the S3 fraction of tetrathionate-grown cells. DEAE cellulose caused an 80 % loss of cytochrome activity. DEAE Sephacel appeared promising at first, with cytochrome being eluted with about 100 mM KCl. However, when the column was re-run with KCl elution of 0-250 mM, all fractions contained similar cytochrome activities, showing no purification had been achieved.

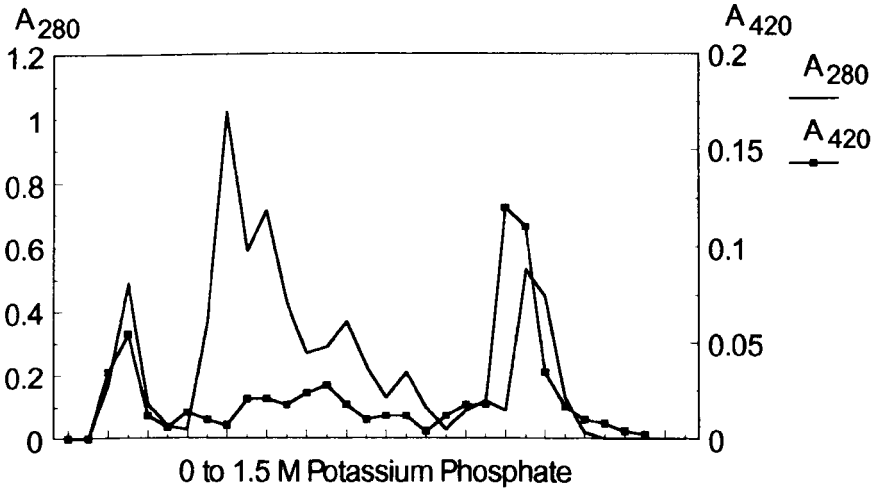
Hydroxylapatite chromatography was used for a purification attempt on the S3 fraction of tetrathionate-grown cells. This gave a small factor of cytochrome purification of 1.5-fold when eluted with 500 mM KH_2PO_4. For this experiment the detergent used was changed to sarkosyl rather than NP 40 as NP 40 is precipitated by phosphate concentrations of above 500 mM. Although this was only a small purification factor, binding and elution of cytochrome to hydroxylapatite had been demonstrated, and full cytochrome activity retained. Therefore, this technique was also used on the oily membrane fraction derived from iron-grown cells. In this case, a 2.5-fold factor of purification was observed with elution of cytochrome occurring at 750 mM KH_2PO_4.

As some reliable, though admittedly small, degree of cytochrome purification had been observed using hydroxylapatite minicolumns, the method was scaled up to 5 ml columns in the hope of increasing the degree of purification. Hydroxylapatite chromatography was first carried out on the S3 fraction of tetrathionate-grown cells, resulting in a 3-fold factor of purification with elution of cytochrome occurring at 500 - 600 mM phosphate. Subsequently hydroxylapatite chromatography was carried out

on the oily membrane fraction of iron-grown cells, resulting in a 4-fold factor of purification with elution of cytochrome occurring at 1 M phosphate and in the flow through material (see Fig 6.5). This was higher than the elution point from the minicolumn, however, the larger column was likely to have produced greater accuracy.

This four-fold purification of 572 nm cytochrome from the oily membrane fraction of iron-grown cells was the largest purification factor attained by column chromatography in this study. Unfortunately insufficient material was available to continue with further purification attempts. However, the final purification factor in comparison with whole cells was twenty-four fold when the six-fold purification factor of membrane preparation is included. This did not provide pure protein as assessed by silver stained SDS-PAGE.

Figure 6.5 Graph illustrating the purification of cytochrome from the oily membrane fraction of iron-grown *Sulfolobus* LM by hydroxylapatite chromatography.



6.7 - Further investigation of the 572 nm cytochrome by electrophoresis

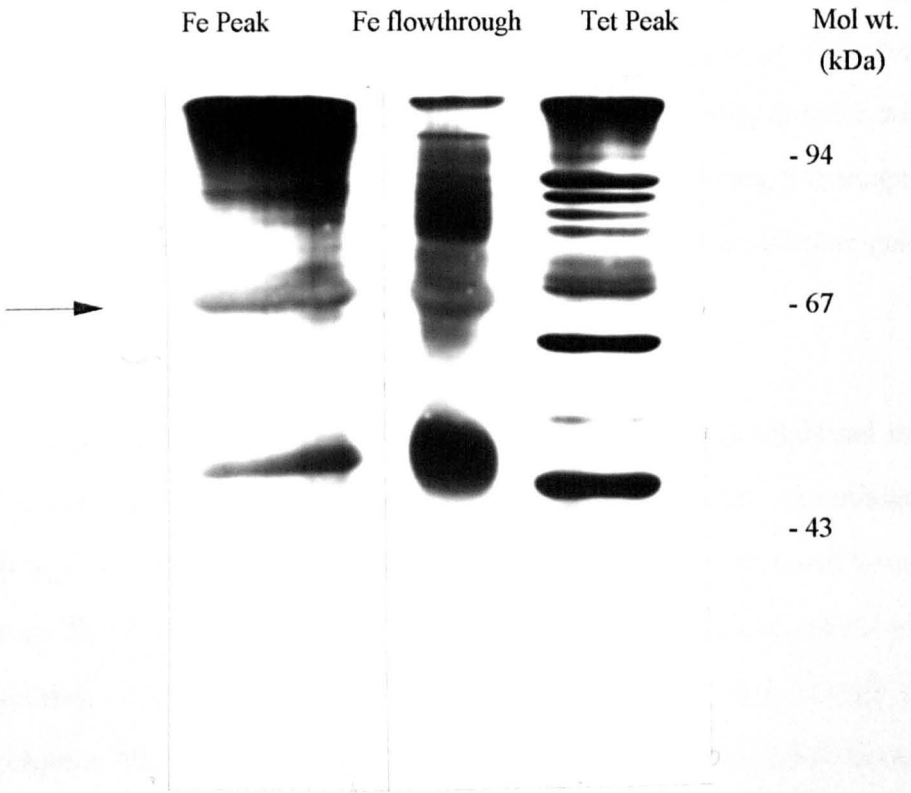
Previous comparisons by SDS-PAGE of chromatography column fractions containing cytochrome activity with those showing no such activity had not lead to the identification of a protein band reproducibly corresponding to the presence of the cytochrome. Similarly, comparison of solublised membrane fraction S3 from iron-grown cells with S3 fraction of tetrathionate-grown cells resulted in no identification of such a protein. Also comparison of solublised membrane fraction S3 from iron-grown cells with S2 fraction of iron-grown cells, containing little cytochrome activity, resulted in similar lack of success. Native PAGE and native PAGE with additional NP 40 was also carried out using solublised membrane fractions, but precipitation of material at the interface between the stacking gel and the resolving gel was seen.

However, with the isolation of the green oily membrane layer from iron-grown cells, a sample with higher specific cytochrome activity was available. SDS-PAGE of this sample versus tetrathionate-grown samples resulted in the identification of protein bands possibly of greater density at approximately 66, 58 and 45 kDa.

After hydroxylapatite chromatography of the oily layer from iron-grown cells, the two fractions with highest specific 572 nm cytochrome activity, the flowthrough material and the fraction eluted by 1 M phosphate, were compared by SDS-PAGE with the sample containing most active 603 nm cytochrome from the hydroxylapatite purification of tetrathionate-grown membrane fraction (see Fig. 6.6). Major protein bands were seen at approximately 40 kDa and 66 kDa in the fractions containing 572 nm cytochrome activity. The 40 kDa band, however, had a counterpart in the fraction derived from tetrathionate-grown cells. The 66 kDa band therefore appeared to be the most likely candidate for the 572 nm cytochrome. It was also observed to give a 'negative' silver stain, whereby a yellow / orange colour was produced rather than the usual black precipitate. This may have been one reason why this band had not

previously been identified. However, this result was not sufficiently conclusive to allow positive identification of this protein band as the 572 nm cytochrome as a number of other minor protein bands were also present.

Figure 6.6 Silver stained SDS-PAGE of fractions (Fe peak and Fe flowthrough) possessing greatest cytochrome activity from hydroxylapatite purification of the oily membrane layer from iron-grown *Sulfolobus* LM and solublised membrane of tetrathionate-grown *Sulfolobus* LM (Tet peak).



In order to provide confirmation of this result native PAGE was carried out on the above samples. However, all gave a similar result to previous native gel attempts with sample being retarded at the interface between the stacking gel and the resolving gel resulting in streaking of proteins on the resolving gel.

Additionally, a sucrose gradient separation of solublised membrane material S3 derived from iron-grown cells was carried out. Cytochrome activity was found in the 40 % and 50 % sucrose fractions. However, again no protein bands specifically related to cytochrome activity could be identified.

TMBZ haem stains had been tried previously using native polyacrylamide gels and native polyacrylamide gels containing extra NP 40 detergent. Generally haem groups are dissociated from cytochromes by SDS sample preparation. However certain c-type cytochromes have been observed to retain haem staining activity with DMB during SDS-PAGE when prepared with SDS sample buffer lacking β-mercaptoethanol and omitting the boiling step and using a semi-denaturing polyacrylamide gel (Francis and Becker, 1984).

This procedure was tested using tetrathionate-grown S3 solublised membrane sample, which gave a haem stained band at approximately 55 kDa. The procedure was repeated with the oily fraction, the S3 solublised membrane fraction and lysed cells of iron-grown *Sulfolobus* LM. These all gave haem staining bands at approximately 66 kDa, agreeing with the position of the negatively stained band derived from the hydroxylapatite purification. In all cases the presence of a barely visible haem staining band running at approximately 55 kDa was noted, corresponding to the band seen by haem staining of tetrathionate-grown material. Additionally both iron-grown and tetrathionate-grown samples were subjected to the usual SDS sample preparation including β-mercaptoethanol and boiling for 5 min. This resulted in the loss of the 55 kDa band but increased resolution of the 66 kDa band (see Fig 6.7). Therefore the

presence of a haem staining protein running at 66 kDa on SDS-PAGE present only in iron-grown cells and in column fractions possessing high 572 nm cytochrome activity was demonstrated. It therefore seems reasonable to assume that this represented the 572 nm cytochrome. This remarkable stability may be due to the thermostability of the protein, as thermostable proteins are often observed to possess considerable stability in the presence of detergents.

Figure 6.7 DMB haem stained SDS-PAGE of membrane pellet P3 of iron-grown (Fe) and tetrathionate-grown (Tet) *Sulfolobus* LM

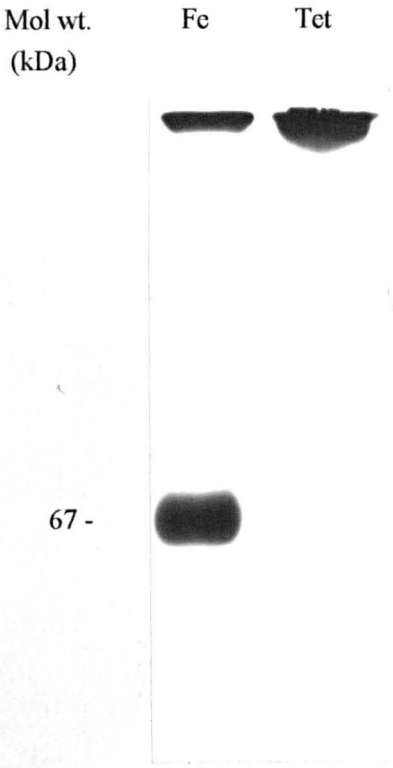
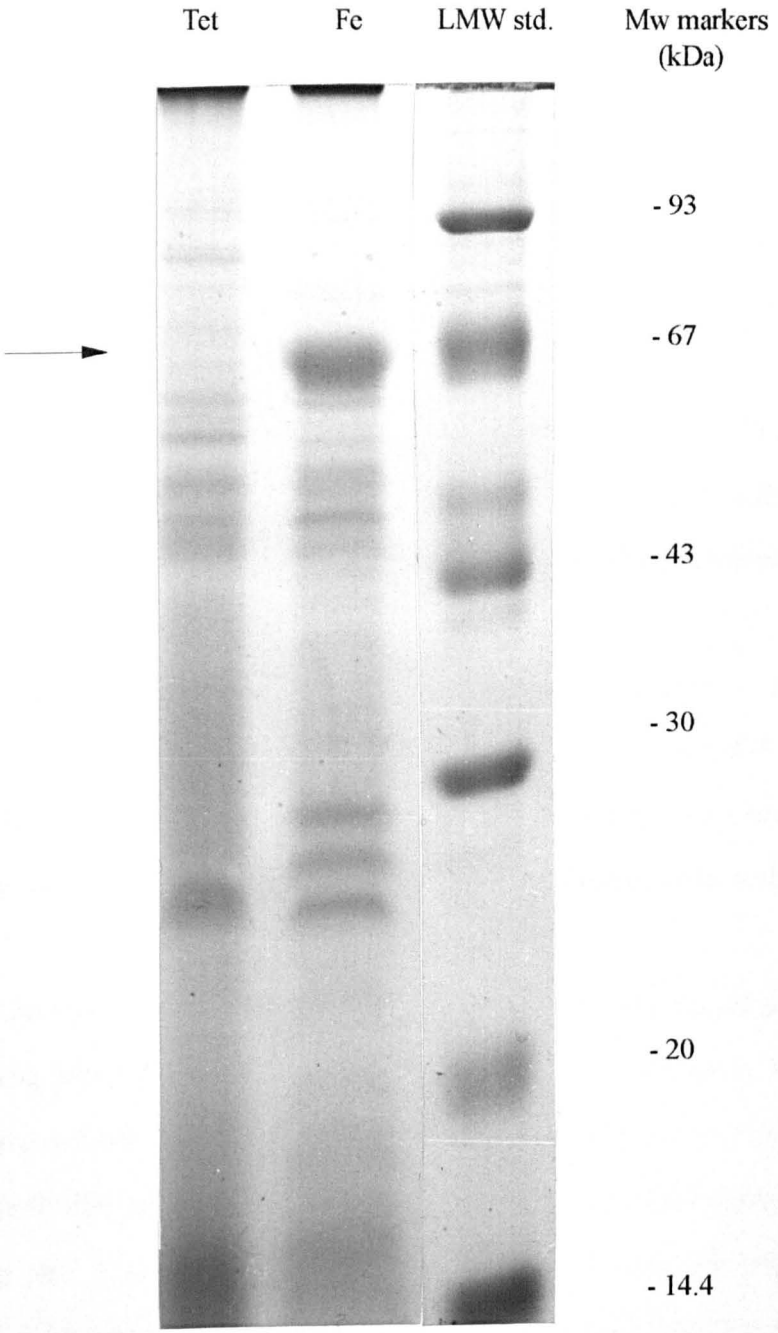


Figure 6.8 Coomassie stained SDS-PAGE of iron-grown membrane pellet P3 (Fe) and tetrathionate-grown membrane pellet P3 (Tet). The 66 kDa band is indicated.



6.8 - Attempted isolation of 66 kDa cytochrome

The 572 nm cytochrome had possibly been identified as running at 66 kDa on SDS-PAGE (see Fig 6.8). The next objective was to isolate semi-purified material for N-terminal protein microsequencing. However, insufficient iron-grown biomass was available for repetition of the hydroxylapatite purification and positive samples of the initial experiment were exhausted. The samples of iron-grown membrane pellet did not show sufficient purity of the 66 kDa protein to justify using this material directly for N-terminal microsequencing. One additional separation step was required.

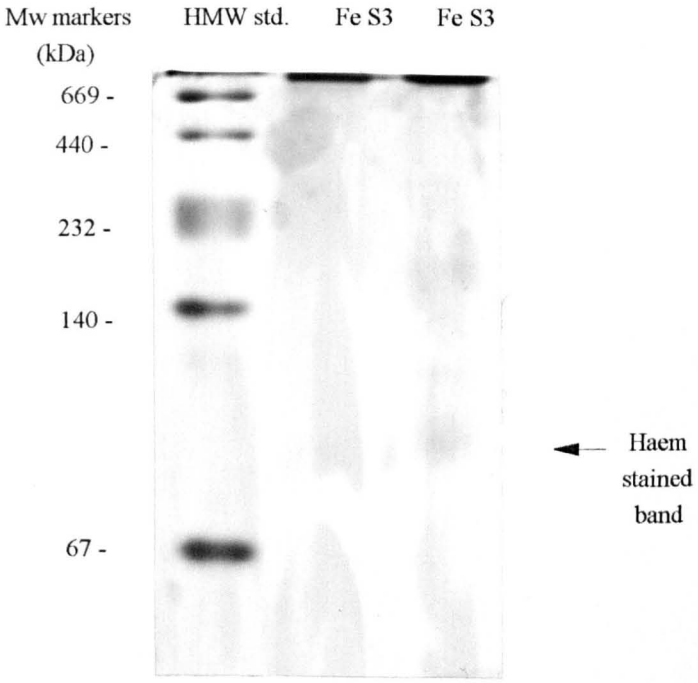
Previously membrane protein samples were seen to be retarded at the interface between the stacking gel and the resolving gel during native PAGE. This was thought to be due to the association of lipids and detergents with the proteins.

The initial attempt to overcome this problem used a slab gel isoelectric focussing method which allowed samples excised from SDS-PAGE to be used, thus in effect allowing two dimensional electrophoresis. However, on Coomassie staining, no discrete band was seen after isoelectric focussing of the excised 66 kDa band.

An final attempt to resolve the cytochrome on a native gel used solublisation of membrane pellet P2 in SB II with the addition of the chaotropes KSCN and DMSO, producing soluble fraction S3. This resulted in the resolution of certain protein bands by native-PAGE, although the majority of proteins were still retarded at the base of the stacking gel. A weakly staining band running at 90 kDa was highlighted by haem staining. This was shown to represent the 66 kDa cytochrome by excision and use on SDS-PAGE. The remaining iron-grown material was amalgamated and used for a purification attempt by running a native PAGE, excising the 90 kDa band and using this material as sample for SDS-PAGE. However, insufficient protein was produced to

allow N-terminal protein microsequencing due to the lack of time to grow more biomass using ferrous iron.

Figure 6.9 Native PAGE of solublised membrane fraction S3 derived from iron-grown cells (Fe S3). The haem staining band is indicated.



In summary, the novel 572 nm cytochrome was found to be present in *Sulfolobus* LM only during ferrous iron oxidation and was shown to be stable in both acidic and neutral conditions. Ultracentrifugation studies of iron-grown *Sulfolobus* LM revealed that the cytochrome was concentrated in the membrane fraction. Partial purification of the cytochrome from solublised membrane fraction suggested that it was represented by a protein band running at 66 kDa on SDS-PAGE. Haem staining supported this and highlighted the stability of the cytochrome. Attempts to purify a sufficient quantity of the cytochrome for N-terminal protein microsequencing were initiated. This objective was not realised due to constraints on the time available to produce more biomass. However, a route to this objective was identified, which may facilitate future studies of this cytochrome.

CHAPTER 7

Summary

This work has initiated a molecular biological approach to the study of two aspects of lithotrophy in thermoacidophilic iron-oxidizing archaea; the oxidation of iron and carbon dioxide assimilation.

Two strains of thermoacidophilic archaea were used in this study, strain HT and strain LM. It was intended to use 16S rRNA sequence data to reliably identify these strains and their relationship to other *Sulfolobus*-like organisms. The gene encoding the 16S rRNA of the novel strain HT was cloned and its entire nucleotide sequence obtained. Additionally, part of the 5' flanking region of the 16S rRNA gene, and part of the intergenic region, up to the expected position of the 23S rRNA precursor processing stem, was sequenced. The clone obtained was expected to contain the 23S rRNA gene and its 3' flanking region, which may provide additional phylogenetic and rRNA processing data if sequenced in future.

The predicted secondary structure of the flanking sequences was similar to that of *Sulfolobus acidocaldarius*, implying a similar mechanism of rRNA transcript processing. Using the 16S rRNA sequence, preliminary phylogenetic analysis was made with reference to *S. acidocaldarius*, *S. shibatae*, *Desulfurococcus mobilis* and *Pyrodictium occultum*. This analysis showed the sequence from strain HT segregating within the genus *Sulfolobus*, and being more similar to the sulphur-oxidizing (but not iron-oxidizing) *S. shibatae* than to the heterotrophic *S. acidocaldarius*. This work was considered sufficiently important to be extended into a larger effort within the research group, and so was not pursued further in this project for strain LM. The relatedness of the *Sulfolobus* HT and LM strains has subsequently been established in a more extensive study of phylogeny amongst thermoacidophilic archaea. This is illustrated in Appendix 1 (kindly provided by Dr. Sara Waterhouse, Biological Sciences, Warwick University).

The components and mechanisms of carbon dioxide fixation during autotrophic growth in the iron-oxidizing crenarcheota have not been characterised. Previous work with *Sulfolobus* strain LM identified a protein of 330 kD, comprised of 59 kDa and 19 kDa subunits, over-expressed during CO_2 limited growth. This was implicated in CO_2 fixation in the presence of ATP and acetyl-CoA (Norris *et al.*, 1989).

The protein was investigated during this study, using a molecular biological approach to its characterisation. The 59 kDa subunit was first partially purified and its N-terminal amino acid sequence obtained. A DNA probe was designed using this data, allowing the cloning and sequencing of the gene encoding this 59 kDa subunit. An open reading frame was found to be adjacent to the gene encoding the 59 kDa subunit. This was predicted to encode a protein of approximately 17 kDa molecular weight. The two genes were arranged in an operon structure, with consensus archaeal transcription initiation and termination sites found upstream and downstream respectively. It appeared likely that these two genes were cotranscribed and that the unknown ORF encoded the subunit previously estimated as 19 kDa by SDS-PAGE.

Homology searches revealed that the predicted amino acid sequence of the 59 kDa subunit was similar to those of ATP dependent biotin carboxylases. The predicted amino acid sequence of the ORF likely to encode the 19 kDa subunit was found to be homologous to those of biotin carboxyl carrier proteins, examples of which have been previously observed to give anomalously high estimated molecular weights by SDS-PAGE analysis.

The predicted amino acid sequence of the 59 kDa protein contained a conserved consensus ATP binding site similar to those of biotin carboxylases and carbamoyl phosphate synthetases. It also contained conserved regions postulated to represent the active site and the bicarbonate binding site in other biotin carboxylases.

The predicted sequence of the C-terminal region of the 19 kDa subunit was homologous to those of biotin carboxyl carrier proteins. This contained a conserved biotin binding motif and conserved residues thought to increase the mobility of this portion of the protein. The positioning of these residues showed similarity to lipoamide dehydrogenase. The N-terminal region of this protein did not show a similar degree of homology, but this region is variable amongst biotin carboxyl carrier proteins.

The two subunits appear to form part of an ATP-dependent carboxylase. However, the additional carboxyl transferase subunits, responsible in other carboxylases for transferring the carboxyl unit from carboxy-biotin to the substrate, have not yet been identified. From previous experiments, it is likely that this enzyme is an acetyl-CoA carboxylase, but it has been shown to be very highly expressed in CO_2-limited cells and to be associated with the major CO_2 assimilation mechanism in these cells. It is possible that this enzyme has a role in CO_2 fixation in the thermoacidophilic archaea, perhaps related to a modified reverse TCA cycle. Future work may allow the further elucidation of this involvement *via* further biochemical characterisation of CO_2 fixation and the study of the carboxyl transferase subunits thought to be required for the activity of this enzyme. It also remains to be determined if this protein is similarly over-expressed during CO_2-limited autotrophic growth in the other archaeal thermoacidophiles such as *Sulfolobus* strain HT.

An investigation of components induced during iron oxidation was initiated by the comparison of iron-grown and tetrathionate-grown *Sulfolobus* strain LM. The major protein apparently produced specifically by *Sulfolobus* strain LM when grown on ferrous iron was identified as a 27 kDa polypeptide by SDS-PAGE. This protein was partially purified and its N-terminal amino acid sequence obtained. Probes were designed using this data, allowing the cloning and sequencing, by a co-worker, of the gene encoding this protein.

Homology searches allowed the identification of the 27 kDa protein as an alkyl hydroperoxide reductase / thiol-specific antioxidant (AhpC/TSA) protein. This was surprising as it could not obviously be assigned a role in the process of ferrous iron oxidation. Nevertheless, the protein may have a vital role in the growth of *Sulfolobus* strain LM in the presence of iron. The major possibilities for such a role are suggested as protection from reactive sulphur species generated by the interaction of ferric iron, oxygen and thiols, or protection from hydroperoxides produced by superoxide dismutase.

The functions of many of the AhpC/TSA proteins investigated in other organisms are unclear, although a general role of protection against oxidative stress appears to be common. This may be further investigated in *Sulfolobus* strain LM by induction studies to determine under which conditions the protein is expressed and, if possible, by the study of non-producing mutants. The protein does appear to be similarly expressed in *Sulfolobus* strain HT, but not expressed to a similar extent in *Acidianus brierleyi*, a difference upon which future work may focus.

The other major protein targeted as being potentially involved in ferrous iron oxidation, an apparently novel cytochrome, was more difficult to study since the protein had to be identified from other whole cell proteins rather than, as with the AhpC/TSA protein, simply selected as obviously over-produced with the switch from growth on a sulphur compound to growth on ferrous iron.

This cytochrome showed an alpha absorbance peak at 572 nm, and was found to be strongly membrane associated. It also showed high stability to changes in pH, retaining redox activity at both pH 2 and pH 7, and showed stability with respect to denaturation by SDS and β-mercaptoethanol treatment, retaining its haem staining activity even after SDS-PAGE. This allowed its identification as a protein of 66 kDa estimated molecular weight. Progress was made in the partial purification of this

protein, identifying a procedure which may allow its isolation in future when more biomass has been produced. This may allow the N-terminal sequencing of this protein and ultimately its molecular biological characterisation.

Difference spectra only of iron-grown cells of other archaeal thermoacidophilic iron oxidizing archaea, *Sulfolobus* HT, *Metallosphaera sedula* and *Acidianus brierleyi*, have previously shown the presence of this cytochrome, supporting the hypothesis that it has a key role in electron transport during iron oxidation. Difference spectra of iron-grown *Sulfolobus* LM also showed an alpha peak at 603 nm, indicative of the presence of an aa_3 type terminal oxidase as found in heterotrophically grown *S. acidocaldarius*. Whether this interacts with the novel cytochrome during iron oxidation remains to be determined.

In summary, progress has been made in a number of aspects of the study of lithoautotrophic growth of the thermoacidophilic archaea. The phylogeny of this group of organisms has been extended by the addition of the first crenarcheotal iron-oxidising species, *Sulfolobus* HT, to the 16S rRNA database. This work has subsequently been expanded into a more comprehensive study of the phylogeny of thermoacidophilic archaea. The putative biotin carboxylase overproduced during growth under CO_2 limitation has been cloned and sequenced. The further biochemical study of this protein may allow elucidation of the as yet uncharacterised CO_2 fixation pathway in *Sulfolobus*. The study of a polypeptide overproduced during iron oxidation, the AhpC/TSA protein, has initiated the molecular study of oxidative stress tolerance in iron-oxidising archaea. In addition, initial stages in the isolation and characterisation of a novel cytochrome present only during iron oxidation have been carried out. Further studies of this cytochrome may allow an insight into the iron oxidation system of *Sulfolobus*.

REFERENCES

AALEN, R.B., OPSAHL-FERSTAD, H.-G., LINNESTAD, C., and OLSEN, O.-A. (1994). Transcripts encoding an oleosin and a dormancy-related protein are present in both the aleurone layer and the embryo of developing barley (*Hordeum vulgare* L.) seeds. Plant J. 5: 385-396.

AL-FEEL, W., CHIRALA, S.S., and WAKIL, S.J. (1992). Cloning of the yeast *FAS3* gene and primary structure of yeast acetyl-CoA carboxylase. Proc. Natl. Acad. Sci. USA. 89: 4534-4538.

ALI, S.T., and GUEST, J.R. (1990). Isolation and characterization of lipoylated and unlipoylated domains of the E2p subunit of the pyruvate dehydrogenase complex of *Escherichia coli*. Biochem. J. 271: 139-145.

AMES, G.F-L., and NIKAIDO, K. (1976). Two dimensional gel electrophoresis of membrane proteins. Biochemistry 15: 616-623.

ANEMÜLLER, S., and SCHÄFER, G. (1989). Cytochrome aa_3 from the thermoacidophilic archaebacterium *Sulfolobus acidocaldarius*. FEBS. Lett. 244: 451-455

ANDREWS, S.C., HARRISON, P.M., and GUEST, J.R. (1991). A molecular analysis of the 53.3 minute region of the *Escherichia coli* linkage map. J. Gen. Microbiol. 137: 361-367.

ARMSTRONG-BUISSERET, L., COLE, M.B., and STEWART, G.S.A.B. (1995). A homologue to the *Escherichia coli* alkyl hydroperoxide reductase AhpC is induced by osmotic shock in *Staphylococcus aureus*. Microbiology 141: 1655-1661.

BARR, D.W., INGLEDEW, E.J., and NORRIS, P.R. (1990). Respiratory chain components of iron-oxidizing, acidophilic bacteria. FEMS Microbiol. Lett. 70: 85-95.

BENNER, S.A., ELLINGTON, A.D., and IAUER, A. (1989). Modern metabolism as a palimpsest of the RNA world. Proc. Natl. Acad. Sci. USA. 86: 7054-7058

BEST, E.A., and KNAUF, V.C. (1993). Organisation and nucleotide sequences of the genes encoding the biotin carboxylase protein of *Pseudomonas aeruginosa* acetyl-CoA carboxylase. J. Bacteriol. 175: 6881-6889.

BLAKE, R., SHUTE, E.A., GREENWOOD, M.M., SPENCER, G.H., and INGLEDEW, W.J. (1993). Enzymes of aerobic respiration on iron. FEMS Microbiol. Rev. 11: 9-18.

BRADFORD, M.M.,(1976). A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72: 248-254.

BRIERLEY, J.A., and BRIERLEY, C.L. (1986). Microbial mining using thermophilic bacteria. *In*, Thermophiles: general, molecular and applied microbiology. (ed. T.D. Brock), pp. 279-305. Wiley, New York.

BROWN, J.W., DANIELS, C.J., and REEVE, J.N. (1989). Gene structure, organisation and expression in archaebacteria. CRC. Crit. Rev. Microbiol. 16: 287-338

BROWNER, M.F., TARONI, F., SZTUL, E., and ROSENBERG, L.E. (1989). Sequence analysis, biogenesis, and mitochondrial import of the α-subunits of rat liver propionyl-CoA carboxylase. J. Biol. Chem. 261: 12680-12685.

CASTREANA, J., LÜBBEN M., and SARASTE, M. (1995). New archaebacterial genes coding for redox proteins: Implications for the evolution of aerobic metabolism. J. Mol. Biol. 250: 202-210.

CHAE, H.Z., CHUNG, S.J., and RHEE, S.G. (1994a). Thioredoxin-dependent peroxide reductase from yeast. J. Biol. Chem. 269: 27670-27678.

CHAE, H.Z., KIM, I.-H., KIM, K., and RHEE, S.G. (1993). Cloning, sequencing, and mutation of thiol-specific antioxidant gene of *Saccharomyces cerevisiae*. J. Biol. Chem. 268: 16815-16821.

CHAE, H.Z., ROBISON, K., POOLE, L.B., CHURCH, G., STORZ, G., and RHEE, S.G. (1994b). Cloning and sequencing of thiol-specific antioxidant from mammalian brain: alkyl hydroperoxide reductase and thiol-specific antioxidant define a large family of antioxidant enzymes. Proc. Natl. Acad. Sci. USA 91: 7017-7021.

COLMER, A.R., TEMPLE, K.T., and HINKLE, M.E. (1950). An iron-oxidising bacterium from the acid mine drainage of some bituminous coal mines. J. Bacteriol. 59: 317-328.

COX, J.C., and BOXER, D.H. (1978). The purification and some properties of rusticyanin, a blue copper protein involved in iron(II) oxidation from *Thiobacillus ferrooxidans*. Biochem. J. 174: 497-502.

CUBELLIS, M., ROZZO, V., NITTI, C., AMONE, G., MARINO M.I., and SANNIA, G. (1989). Cloning and sequencing of the gene coding for aspartate aminotransferase from the thermophilic archaebacterium, *Sulfolobus solfataricus*. J. Biochem. 186: 375-381.

DANSON, M.J. (1989). Central metabolism of the archaebacteria - an overview. Can. J. Microbiol. 35: 59-64

DANSON, M.J., and HOUGH, D.W. (1991). The enzymology of archaebacterial pathways of central metabolism. Biochem. Soc. Symp. 58: 7-21

DUROVIC, P., and DENNIS, P.P. (1994). Separate pathways for excision and processing of 16S rRNA from the primary rRNA operon transcript from the hyperthermophilic archaebacterium, *Sulfolobus acidocaldarius* : Similarities to eukaryotic rRNA processing. Mol. Microbiol. 13: 229-242

FALL, R.R., ALBERTS, A.W., and VAGELOS, P.R. (1975). Analysis of bacterial biotin proteins. Biochim. Biophys. Acta 379: 496-503

FELSENSTEIN, J. (1995). PHYLIP (Phylogeny Inference Package), Version 3.57c. Department of Genetics, University of Washington, Seattle.

FERGUSON, S.J. (1988). Periplasmic electron transport reactions. *In* Bacterial Energy Transduction, (ed. C. Anthony). Academic Press, U.K. : 151-180

FOX, G.E., STECKEBRANDT, E., HESPELL, R.B.M., GIBSON, J., MANILOFF, J., DYER, T.A., WOLFE, R.S., BALCH, W.E., TANNER, R., MAGRUM, L.J., ZABLEN, L.B., BLAKEMORE, R., GUPTA, R., BONEN, L., LEWIS, B.J., STAHL, D.L., LUEHRSEN, K.R., CHEN, K.N., and WOESE, C.R. (1980). The phylogeny of prokaryotes. Science 209: 457-463.

FRANCIS, R.T., and BECKER, R.B. (1984). Specific indication of hemoproteins in polyacrylamide gels using a double-staining process. Anal. Biochem. 136: 509-514

FUCHS, G. (1988). Alternative pathways of autotrophic CO_2 fixation. *In* Autotrophic bacteria (ed. H.G. Schlegel and B. Bowien), pp. 365-382. Springer Verlag.

FUCHS, G., and STUPPERICH, E. (1986). Carbon assimilation pathways in archaebacteria. *In* Archaebacteria '85 (ed. O. Kandler and W. Zillig), pp. 364-369. Gustav Fischer Verlag, Stuttgart.

FUKUMORI, Y., YANO, T., SATO, A., and YAMANAKA, T. (1988). Fe(II)-oxidising enzyme purified from *Thiobacillus ferrooxidans*. FEMS. Microbiol. Lett. 50: 169-172.

GENBAUFFE, F.S., and COOPER, T.G. (1991). The urea amidolyase (DUR1,2) gene of *Saccharomyces cerevisiae*. DNA Seq. 2: 19-32.

GOLDMARK, P.J., CURRY, J., MORRIS, C.F., and WALKER-SIMMONS, M.K. (1992). Cloning and expression of an embryo-specific mRNA up-regulated in hydrated dormant seeds. Plant Mol. Biol. 19: 433-441.

GOODHEW, C.F., BROWN, K.R., and PETTIGREW, G.W. (1986). Haem staining in gels, a useful tool in the study of bacterial c-type cytochromes. Biochim. et Biophys. Acta 852: 288-294.

GORNICKI, P., SCAPPINO, L.A., and HASELKORN, R. (1993). Genes for two subunits of acetyl coenzyme A carboxylase of *Anabaena* sp. strain PCC 7120: Biotin carboxylase and biotin carboxyl carrier protein. J. Bacteriol. 175: 5268-5272.

GROGAN, D.W. (1989). Phenotypic characterization of the archaebacterial genus *Sulfolobus*: comparison of five wild-type strains. J. Bacteriol. 171: 6710-6719.

GROGAN, D.W., PALM, P., and ZILLIG, W. (1990). Isolate B12 which harbours a virus-like element, represents a new species of the archaebacterial genus *Sulfolobus*, *Sulfolobus shibatae*, sp. nov. Arch. Microbiol. 154: 594-599.

HART, A., MURRELL, J.C., POOLE, R.K., and NORRIS, P.R. (1991). An acid-stable cytochrome in iron-oxidising *Leptospirillum ferrooxidans*. FEMS. Microbiol. Lett. 81: 89-94.

HENRIKSON, K.P., and ALLEN, S.H.G. (1979). Purification and subunit structure of propionyl coenzyme A carboxylase of *Mycobacterium smegmatis*. J. Biol. Chem. 254: 5888-5891.

HENSEL, R. (1993). Proteins of extreme thermophiles. *In* The biochemistry of Archaea (Archaebacteria). (ed. M. Kates, D. J. Kushner, and A. T. Matheson), pp. 209-221. Elsevier, Amsterdam.

HUBER, G., and STETTER, K.O. (1991). *Sulfolobus metallicus*, sp. nov., a novel strictly chemolithoautotrophic thermophilic archaeal species of metal-mobilizers. Syst. Appl. Microbiol. 14: 372-378.

HUDSON-TAYLOR, D.E., DOLAN, S.A., KLOTZ, F.W., FUJIOKA, H., AIKAWA, M., KOONIN, E.V., and MILLER, L.H. (1995). *Plasmodium falciparum* protein associated with the invasion junction contains a conserved oxidoreductase domain. Mol. Microbiol. 15: 463-471.

HUNAITI, A.R., and KOLATTUKUDY, P.E. (1982). Isolation and characterisation of an acyl-coenzyme A carboxylase from an erythromycin-producing *Streptomyces erythreus*. Arch. Biochem. Biophys. 216: 362-371

HUTCHINS, S.R., DAVIDSON, M.S., BRIERLEY, J.A., and BRIERLEY, C.L. (1986). Microorganisms in reclamation of metals. Ann. Rev. Microbiol. 40: 311-336

HUTCHINS, S.R., BRIERLEY, J.A., and BRIERLEY, C.L. (1987). Microbial pretreatment of refractory sulfide and carbonaceous gold ores. *In* Process Mineralogy VII : Applications to mineral benefication technology and mineral exploitation, with special emphasis on disseminated carbonaceous gold ores (ed. A.H. Vassiliou, D.M. Hausen and D.J.T. Carson), pp. 53-56. The Metallurgical Soc. Inc., Warrendale, Pennsylvania.

INGLEDEW, W.J. (1982). *Thiobacillus ferrooxidans* - the bioenergetics of an acidophilic chemolithotroph. Biochim. et Biophys. Acta 683: 89-117

ISHII, T., YAMADA, M., SATO, H., MATSUE, H., TAKETANI, S., NAKAYAMA, K., SUGITA, Y., and BANNAI, S. (1993). Cloning and characterization of a 23-kDa stress-induced mouse peritoneal macrophage protein. J. Biol. Chem. 268: 18633-18636.

JACOBSON, F.S., MORGAN, R.W., CHRISTMAN, M.F., and AMES, B.N. (1989). An alkyl hydroperoxide reductase from *Salmonella typhimurium* involved in the defense of DNA against oxidative damage. J. Biol. Chem. 264:1488-1496.

KANDLER, O., and STETTER, K.O. (1981). Evidence for autotrophic CO_2 assimilation in *Sulfolobus brierleyi* via a reductive carboxylic acid pathway. Zbl. Bakt. Hyg., I. Abt. Orig. C., 2: 111-121.

KARAVAIKO, G.I., GOLOVACHEVA, R.S., PIVOVAROVA, T.A., TZAPLINA, I.A., and VARTANJAN, N.S. (1988). Thermophilic bacteria of the genus *Sulfobacillus*. *In* Biohydrometallurgy 1988 (ed. P.R. Norris and D.P. Kelly), pp. 29-41. Science and Technology Letters, Kew, U.K.

KELLY, D.P. (1988) Evolution of the understanding of the microbiology and biochemistry of the mineral leaching habitat. *In* Biohydrometallurgy 1988 (ed. P.R. Norris and D.P. Kelly), pp. 3-14. Science and Tecnology Letters, Kew, U.K.

KIM, I.H., KIM, K., and RHEE, S.G. (1989). Induction of an antioxidant protein of *Saccharomyces cerevisiae* by O_2, Fe^{3+}, or mercaptoethanol. Proc. Natl. Acad. Sci. USA 86: 6018-6022.

KIM, K., KIM, I.L., LEE, K.-Y., RHEE, S.G. and STADTMAN, E.R. (1988). The isolation and purification of a specific "protector" protein which inhibits enzyme inactivation by a thiol/Fe(III)/O_2 mixed-function oxidation system. J. Biol. Chem. 263: 4704-4711.

KLEITZIN, A. (1992). Molecular characterisation of the *sor* gene which encodes the sulfur oxygenase / reductase of the thermoacidophilic archaeum *Desulfurolobus ambivalens*. J. Bact. 174: 5854-5859

KLENK, H.P., and DOOLITTLE, W.F. (1994). Archaea and eukaryotes versus bacteria ? Curr. Biol. 4: 920-922

KNOWLES, J.R. (1989). The mechanism of biotin-dependent enzymes. Ann. Rev. Biochem. 58: 195-221.

KONDO, H., SHIRATSUCHI, K., YOSHIMOTO, T., MASUDA, T., KITAZONO, A., TSURU, D., ANAI, M., SEKIGUCHI, M., and TANABE, T. (1991). Acetyl-CoA carboxylase from *Escherichia coli*: Gene organisation and nucleotide sequence of the biotin carboxylase subunit. Proc. Natl. Acad. Sci. USA. 88: 9730-9733.

KUROSAWA, N., and ITOH, Y.M. (1993). Nucleotide sequence of the 16S rRNA gene from thermophilic archaea *Sulfolobus acidocaldarius* ATCC 33909. Nucleic Acids Res. 21: 357

KUSANO, T., TAKESHIMA, T., SUGAWARA, K., INOUE, C., SHIRATORI, T., YANO, T., FUKUMORI, Y., and YAMANAKA, T. (1992). Molecular cloning of the gene encoding *Thiobacillus ferrooxidans* Fe(II) oxidase. J. Biol. Chem. 267: 11242-11247.

LAMHONWAH, A.-M., MAHURAN, D., and GRAVEL, R.A. (1989). Human mitochondrial propionyl-CoA carboxylase: localisation of the N-terminus of the pro- and mature α chains in the deduced primary sequence of a full length cDNA. Nucleic Acids Res. 17: 4396

LANE, D.J. (1991). 16S / 23S rRNA sequencing. *In* Nucleic acid techniques in bacterial systematics, (ed. E. Stackebrandt and M. Goodfellow). J. Wiley and Sons, Chichester.

LANE, M.D., MOSS, J., and POLAKIS, S.E. (1974). Acetyl coenzyme A carboxylase. Curr. Top. Cell Regul. 8: 139-187

LE ROUX, N.W., and WAKERLEY, D.S. (1987). Leaching of chalcopyrite ($CuFeS_2$) at 70°C using *Sulfolobus*. *In* Biohydrometallurgy, (ed. P.R. Norris and D.P. Kelly). pp. 305-317. Science and Technology Letters, Kew.

LI, S.-J., and CRONAN, J.E. (1992) The gene encoding the biotin carboxylase subunit of *Escherichia coli* acetyl-CoA carboxylase. J. Biol. Chem. 267: 855-863.

LIM, Y.-S., CHA, M.-K., KIM, H.-K., and KIM, I.-H. (1994). The thiol-specific antioxidant protein from human brain: gene cloning and analysis of conserved cysteine regions. Gene 140: 279-284.

LIM, F., MORRIS, C.P., OCCHIODORO, F., and WALLACE, J.C. (1988). Sequence and domain structure of yeast pyruvate carboxylase. J. Biol. Chem. 263: 11493-11497.

LIN, X.-l., and TANG, J. (1990). Purification, characterisation and gene cloning of thermopsin, a thermostable acid protease from *Sulfolobus acidocaldarius*. J. Biol. Chem. 265: 1490-1495.

LIU, X., LINDSTRÖM, E.B., and PETERSSON, S. (1991). Bioleaching of refractory arsenical pyrite concentrates to enhance gold extraction. Scandinavian Journal of Metal Extraction 20: 346-350.

LÓPEZ-CASILLAS, F., BAI, D.-H., LUO, X., KONG, I.-S., HERMODSON, M.A., and KIM, K.-H. (1988). Structure of the coding sequence and primary amino acid sequence of acetyl-coenzyme A carboxylase. Proc. Natl. Acad. Sci. USA. 85: 5784-5788.

LOWRY, O.H., ROSEBROUGH, N.J., FARR, A.L., and RANDAL, R.J. (1951). Protein measurement with folin phenol reagent. J. Biol. Chem. 193: 265-275.

LÜBBEN, M., KOLMERER, B., and SARASTE, M. (1992). An archaebacterial terminal oxidase combines core structures of two mitochondrial respiratory complexes. EMBO. J. 11: 805-812

LÜBBEN, M., WARNE, A., ALBRACHT, S.P.J., and SARASTE, M. (1994). The purified Sox ABCD quinol oxidase complex of *Sulfolobus acidocaldarius* contains a novel haem. Mol. Microbiol. 13: 327-325

LUDWIG, W., and SCHLEIFER, K.M. (1994) Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. FEMS. Micro. Revs. 15: 155-173

MARSH, R.M., NORRIS, P.R., and LE ROUX, N.W. (1983). Growth and mineral oxidation studies with *Sulfolobus*. *In* Recent progress in biohydrometallurgy. (ed. G. Rossi, and A. E. Torma), pp. 71-81. Associazione Mineraria Sarda, Iglesias.

MARSH, T.L., REICH, C.I., WHITELOCK, R.B., and OLSEN, G.J. (1994). Trancription factor IID in the archaea : Sequences in the *Thermococcus celer* genome would encode a product closely related to the TATA binding protein in eukaryotes. Proc. Natl. Acad. Sci. USA. 91: 4180-4184

MATHIEU, I., MEYER, J., and MOULIS, J.-M. (1992). Cloning, sequencing and expression in *Escherichia coli* of the rubredoxin gene from *Clostridium pasteurianum*. Biochem. J. 285: 255-262.

MEILE, L., FISCHER, K., and LEISINGER, T. (1995). Characterization of the superoxide dismutase gene and its upstream region from *Methanobacterium thermoautotrophicum* Marburg. FEMS Microbiol. Lett. 128: 247-253.

MORRIS, C.P., LIM, F., and WALLACE, J.C. (1987). Yeast pyruvate carboxylase: gene isolation. Biochem. Biophys. Res. Comm. 145: 390-396.

MURAMATSU, S., and MIZUNO, T. (1989). Nucleotide sequence of the *fabE* gene and flanking regions containing a bent DNA sequence of *Escherichia coli*. Nucleic Acids Res. 17: 3982.

MURTIF, V.L., BAHLER, C.R., and SAMOLS, D. (1985). Cloning and expression of the 1.3S biotin-containing subunit of transcarboxylase. Proc. Natl. Acad. Sci. USA 82: 5617-5621.

NIIMURA, Y., OHNISHI, K., YARITA, Y., HIDAKA, M., MASAKI, H., UCHIMURA, H., SUZUKI., H., KOZAKI, M., and UOZUMI, T. (1993). A flavoprotein functional as NADH oxidase from *Amphibacillus xylanus* Ep01: purification and characterization of the enzyme and structural analysis of its gene. J. Bacteriol. 175: 7945-7950.

NIXON, A., and NORRIS, P.R. (1992). Autotrophic growth and inorganic sulphur compound oxidation by *Sulfolobus* sp. in chemostat culture. Arch. Microbiol. 157: 155-160.

NORRIS, P.R. (1990). Acidophilic bacteria and their activity in mineral sulphide oxidation. *In* Microbial mineral recovery (ed. H.L. Erlich and C.L. Brierley), pp 3-27. McGraw-Hill Publishing Co., New York.

NORRIS, P.R. 1992. Mineral sulphide oxidation by thermoacidophilic acidophiles, abstr. *In* Thermophiles: science and technology, p. 148. Icetec, Reyjavík.

NORRIS, P.R., and BARR, D.W. (1985). Growth and iron-oxidation by moderate thermophiles. FEMS. Microbiol. Lett. 28: 221-224.

NORRIS, P.R., MARSH, R.M., and LINDSTRÖM, E.B. (1986). Growth of mesophilic and thermophilic acidophilic bacteria on sulphur and tetrathionate. Biotech. Appl. Biochem. 8: 318-329

NORRIS, P.R., and OWEN, J.P. (1992). Strain selection for high temperature oxidation of mineral sulfides in reactors, *In* Harnessing biotechnology for the 21st century. (ed. M.R. Ladisch, and A. Bose), pp. 445-448. American Chemical Society, Washington DC.

NORRIS, P.R., NIXON, A., and HART, A. (1989). Acidophilic, mineral-oxidizing bacteria: the utilization of carbon dioxide with particular reference to autotrophy in *Sulfolobus*, *In* Microbiology of extreme environments and its potential for biotechnology. (ed. M.S. de Costa, J.C. Duarte, and R.A.D. Williams), pp. 24-43. Elsevier, London.

NORRIS, P.R. and PARROT, L. (1986). High temperature, mineral concentrate dissolution with *Sulfolobus*. *In* Fundamental and Applied Biohydrometallurgy, (ed. R.W. Lawrence *et al.*). Elsevier, Amsterdam, pp. 355-365.

NYUNOYA, H., BROGLIE, K.E., WIDGREN, E.E., and LUSTY, C.J. (1985). Characterisation and derivation of the gene coding for mitochondrial carbamyl phosphate synthetase I of rat. J. Biol. Chem. 260: 9346-9356.

OLSEN, G., PACE, N., NUELL, M., KLAINE, B., GUPTA, R., and WOESE, C. (1985) Nucleotide sequence of the thermoacidophilic archaebacterium *Sulfolobus solfataricus*. J. Mol. Evol. 22: 301-307

OLSEN, G.J., WOESE, C.R., and OVERBEEK, R. (1994). The winds of (evolutionary) change breathing new life into microbiology. J. Bacteriol. 176: 1-6.

O'TOOLE, P.W., LOGAN, S.M., KOSTRZYNSKA, M., WADSTRÖM, T., and TRUST, T.J. (1991). Isolation and biochemical and molecular analyses of a species-specific antigen from the gastric pathogen *Helicobacter pylori*. J. Bacteriol. 173: 505-513.

POOLE, R.K. (1988). Bacterial cytochrome oxidases. *In* Bacterial Energy Transduction, (ed. C. Anthony). Academic Press, U.K. : 231-270.

POST, L.E., POST, D.J., and RAUSHEL, F.M. (1989). Dissection of the functional domains of *Escherichia coli* carbamoyl phosphate synthase by site directed mutagenesis. J. Biol. Chem. 265: 7742-7747.

RADFORD, S.E., LANE, E.D., PERHAM, R.N., MARTIN, S.R., and APPELLA, E. (1989). Conformational flexibility and folding of synthetic peptides representing an interdomain segment of polypeptide chain in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. J. Biol. Chem. 264: 767-775.

REED, S.L., FLORES, B.M., BATZER, M.A., STEIN, M.A., STROEHER, V.L., CARLTON, J.E., DIEDRICH, D.L., and TORIAN, B.E. (1992). Molecular and cellular characterization of the 29-kilodalton peripheral membrane protein of *Entamoeba histolytica*: differentiation between pathogenic and non-pathogenic isolates. Infect. Immun. 60: 542-549.

REINSTEIN, J., BRUNE, M., and WITTINGHOFER, A. (1988). Mutations in the nucleotide binding loop of adenylate kinase. Biochemistry 27: 4712-4720.

REITER, W.-D., HÜDEPOHL, U., and ZILLIG, W. (1990). Mutational analysis of an archaebacterial promoter: essential role of a TATA box for transcription efficiency and start-site selection *in vitro*. Proc. Natl. Acad. Sci. USA 87: 9509-9513.

REITER, W.-D., PALM, P., and ZILLIG, W. (1988). Transcription termination in the archaebacterium *Sulfolobus*. Signal structures and linkage to transcription initiation. Nucleic Acids Res. 16: 2445-2459.

RUBIN, R.W., and LEONARDI, C.L. (1983). Two dimensional polyacrylamide gel electrophoreis of membrane proteins. Meth. Enzymol. 96: 184-193

SAMBROOK, J., FRITSCH, E.F., and MANIATIS, T. (1989). Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SANGER, F., NICKLESEN, S., and COULSON, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. 74: 5463-5467.

SATO, A., FUKUMORI, Y., YANO, T., KAI, M., and YAMANAKA, T. (1989). *Thiobacillus ferrooxidans* cytochrome c-552 - purification and some of its molecular features. Biochim. et Biophys. Acta 976: 129-134

SAUNDERS, S.E., and BURKE, J.F. (1990). Rapid isolation of miniprep DNA for double strand sequencing. Nucleic Acids Res. 18: 4948

SCHAFER, S., GOTZ, M., EISENREICH, W., BACHER, A., and FUCHS, G. (1989). ^{13}C-NMR study of autotrophic CO_2 fixation in *Thermoproteus neutrophilus*. Eur. J. Biochem. 184: 151-156

SCUDIERO, R., CAPASSO, C., MADONNA, L., NAZZARO, F., MARULLI, F., DE PRISCO, P.P., LA CARA, F., and PARISI, E. (1992). Identification of a metal-binding complex from thermophilic bacteria grown in the presence of heavy metals, abstr., *In* Thermophiles: Science and Technology. p. 38. Icetec, Reykjavík.

SCHWARZ, E., OESTERHELT, D., REINKE, H., BEYREUTHER, K., and DIMROTH, P. (1988). The sodium ion translocating oxaloacetate decarboxylase of *Klebsiella pneumoniae*. J. Biol. Chem. 263: 9640-9645

SHORT, J.M., FERNANDEZ, J.M., SORGE, J.A., and HUSE, W.D. (1988). λ ZAP: a bacteriophage λ expression vector with *in vivo* excision properties. Nucleic Acids Res. 16: 7583-7600.

SONG, J., WURTELE, E.S., and NIKOLAU, B.J. (1994). Molecular cloning and characterisation of the cDNA coding for the biotin-containing subunit of 3-metylcrotonyl-CoA carboxylase: Identification of the biotin carboxylase and biotin-carrier domains. Proc. Natl. Acad. Sci. USA. 91: 5779-5783.

SPENCER, P.A., BUDDEN, J.R., and SNEYD, R. (1989). Use of a moderately thermophilic bacterial culture for the treatment of refractory arsenopyrite concentrate. *In* Biohydrometallurgy 1989 (ed. J. Salley, R.G.L. McCready and P.L. Wichlacz) pp. 231-242. CANMET SP89-10, Ottawa.

STETTER, K. (1986). Diversity of extremely thermophilic archaebacteria. Gen. Mol. Appl. Microbiol. 39-49.

TAI, S.S., and ZHU, Y.Y. (1995). Cloning of a *Corynebacterium diphtheriae* iron-repressible gene that shares sequence homology with the AhpC subunit of alkyl hydroperoxide reductase of *Salmonella typhimurium*. J. Bacteriol. 177: 3512-3517.

TARTAGLIA, L.A., STORZ, G., BRODSKY, M.H., LAI, A., and AMES, B.N. (1990). Alkyl hydroperoxide reductase from *Salmonella typhimurium*. J. Biol. Chem. 265: 10535-10540.

TIKHONOVA, G.V., LISENKOVA, L.L., DOMAN, N.G., and SKULACHEV, V.P. (1987). Electron transport pathways in *Thiobacillus ferrooxidans*. Biokhimiya 32: 599-606

TIPTON, P.A., and CLELAND, W.W. (1988). Catalytic mechanism of biotin carboxylase: Steady-state kinetic investigations. Biochemistry 27: 4317-4325

TOH, H., KONDO, H., and TANABE, T. (1993). Molecular evolution of biotin-dependent carboxylases. Eur. J. Biochem. 215: 687-696.

WAKAGI, T., YAMAUCHI, T., OSHIMA, T., MÜLLER, M., AZZI, A., and SONE, N. (1989). A novel A-type terminal oxidase from *Sulfolobus acidocaldarius* with cytochrome c oxidase activity. Biochem. Biophys. Res. Comm. 165: 1110-1114

WALDROP, G., HOLDEN, H.M., and RAYMENT, I. (1994). Preliminary X-ray crystallographic analysis of biotin carboxylase isolated from *Escherichia coli*. J. Mol. Biol. 235: 367-369.

WATANABE, S., KOHNO, H., KOUYAMA, H., HIROI, T., YAGO, N., and NAKAZAWA, T. (1994). Purification and characterization of a substrate protein for mitochondrial ATP-dependent protease in bovine adrenal-cortex. J. Biochem. 115: 648-654.

WOESE, C.R., and FOX, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA. 74: 5088-5099.

WOEHLKE, G., WIFLING, K., and DIMROTH, P. (1992). Sequence of the sodium iron pump oxaloacetate decarboxylase from *Salmonella typhimurium*. J. Biol. Chem. 267: 22798-22803.

WOOD, A.P., and KELLY, D.P. (1985). Autotrophic and mixotrophic growth and metabolism of some moderately thermophilic iron-oxidising bacteria. *In* Planetary Ecology (ed. D.E. Caldwell, J.A. Brierley and C.L. Brierley), pp. 251-262. Van Nostrand Reinhold, New York.

WOOD, A.P., KELLY, D.P., and NORRIS, P.R. (1987). Autotrophic growth of four *Sulfolobus* strains on tetrathionate and the effect of organic nutrients. Arch. Microbiol. 146: 382-389.

WOOD, H.G., and BARDEN, R.E. (1977). Biotin enzymes. Ann. Rev. Biochem. 46: 385-413.

WRAY, W., BOULIKAS, T., WRAY, V.P., and HANCOCK, R. (1981). Silver staining of proteins in polyacrylamide gels. Anal. Biochem. 118: 197-203.

XU, X., KOYAMA, N., CUI, M., YAMAGISHI, A., NOSOH, Y., and OSHIMA, T. (1991). Nucleotide sequence of the gene encoding NADH dehydrogenase from an alkalophile, *Bacillus* sp. strain YN-1. J. Biochem. 109: 678-683.

YAMAGUCHI, R., MATSUO, K., YAMAZAKI, A., TAKAHASHI, M., FUKASAWA, Y., WADA, M., and ABE, C. (1992). Cloning and expression of the gene for the Avi-3 antigen of *Mycobacterium avium* and mapping of its epitopes. Infect. Immun. 60: 1210-1216.

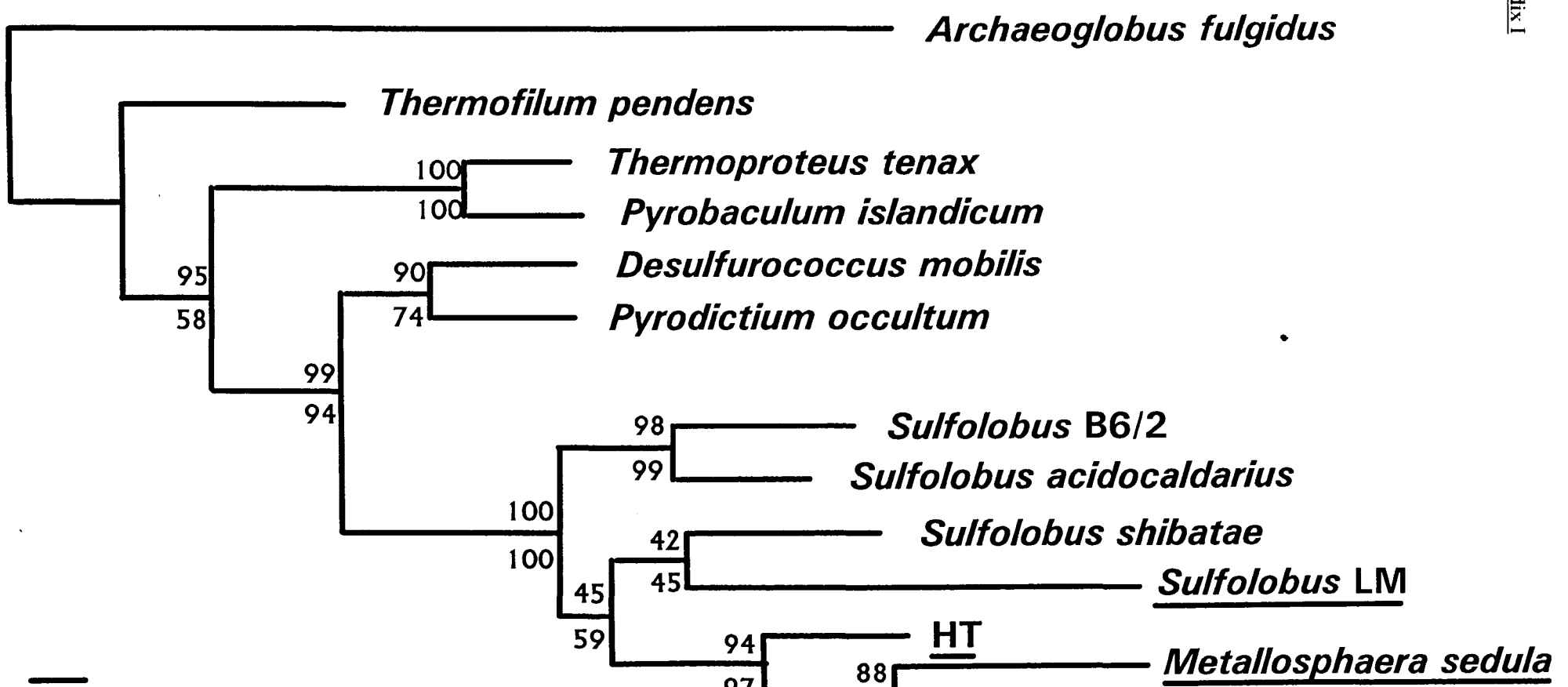
YAMAMOTO, T., MATSUI, Y., NATORI, S., and OBINATA, M. (1989). Cloning of a housekeeping-type gene (*MER5*) preferentially expressed in murine erythroleukemia cells. Gene 80: 337-343.

YAMANAKA, T., KAI, M., YANO, T., and FUKUMORI, Y. (1989). Cytochrome oxidase of an acidophilic iron oxidising bacterium, *Thiobacillus ferrooxidans*, functions at pH 3.5. Biochem. Biophys Res. Comm. 160: 839-843.

ZHANG, J., XIA, W.-L., BREW, K., and AHMAD, F. (1993). Adipose pyruvate carboxylase : amino acid sequence and domain structure deduced from cDNA sequencing. Proc. Natl. Acad. Sci. USA. 90: 1766-1770.

ZILLIG, W. (1993). Confusion in the assignments of *Sulfolobus* sequences to *Sulfolobus* species. Nucleic Acids Res. 21: 5273

APPENDIX



1% substitution

Iron-oxidising species are underlined.